Unmixing the Music: Sparse, Low-Rank, and Deep Compressed-Sensing Approaches to Music Source Separation

Taka Khoo [1]

Abstract

Music source separation, recovering vocals and accompaniment from a single mixture, is approached either by classical signal-processing decompositions or by large end-to-end networks. We show that a wide span of these methods are one idea: recovery of a structured (sparse or low-rank) representation of the time-frequency spectrogram. On a shared short-time Fourier front end we build four separators, Robust PCA, supervised non-negative matrix factorization, K-SVD with sparse-representation classification, and a deep sparse coding network that unrolls a non-negative proximal solver and learns its dictionaries end-to-end, and place them on a single structure-recovery spectrum. On 25 held-out MUSDB18 tracks the methods order cleanly: training-free baselines trail supervised dictionaries, and the unrolled deep coder is the strongest non-oracle separator on both sources (vocal SI-SDR 2.0 dB, accompaniment 8.3 dB) while being the fastest learned model. We pair every method with the recovery condition that governs it and report a negative result: the learned dictionaries are highly coherent ($\mu \approx 0.85$), so worst-case guarantees are vacuous at the operating sparsity, yet separation succeeds. Code, models, and audio are released.

1. Introduction

Single-channel music source separation recovers the constituent stems of a recording given only the final mixture. It is hard because mixing destroys information: many stem configurations explain the same mixture (Cano et al., 2019). Progress has come from two largely separate traditions. The first is model-based: non-negative matrix factorization (NMF) (Lee & Seung, 1999; Smaragdis, 2007), low-rank-plus-sparse decompositions (Huang et al., 2012), and dictionary learning (Aharon et al., 2006) impose explicit structural priors. The second is data-driven: deep networks trained end-to-end (Stöter et al., 2019; Défossez et al., 2021; Rouard et al., 2023) define the current state of the art.

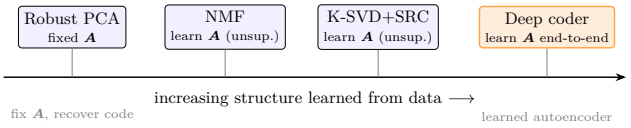

Figure 1. The four separators as points on one structure-recovery spectrum. All solve the same underdetermined system $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ with a sparse code; they differ only in how much of the dictionary $\boldsymbol{A}$ is learned, from a fixed convex split to a fully trained unrolled solver.

The unifying view. These traditions are two ends of a single axis (Figure 1). The magnitude spectrogram of a voiced frame is sparse (a few formant atoms) and the sustained accompaniment is low-rank. These are not separate assumptions but two sides of the same coin: the rank of a matrix is the ℓ_0 norm of its singular spectrum and the nuclear norm is the corresponding ℓ_1 relaxation, so "few active atoms" and "few active singular vectors" are the same prior applied to a vector and to a matrix (Candès et al., 2011). The recovery problem is in every case the underdetermined system $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ with a sparse code: the classical half of the subject fixes the dictionary $\boldsymbol{A}$ and recovers the code, while dictionary learning, K-SVD, and NMF instead learn $\boldsymbol{A}$ from data. Deep masking networks sit at the learned end of this same continuum: a sparse coding network is, in a precise sense, a learned autoencoder (Gregor & LeCun, 2010; Papyan et al., 2017), an unrolled sparse-recovery solver whose operators are trained by backpropagation (Monga et al., 2021).

[1]Thayer School of Engineering, Dartmouth College, Hanover, NH, USA. Correspondence to: Taka Khoo <taka.khoo.th@dartmouth.edu>.

Proceedings of the 42^{nd} International Conference on Machine Learning, Vancouver, Canada. PMLR 267, 2025.

Contributions. (i) We cast vocal/accompaniment separation as structured recovery and instantiate four separators that share a short-time Fourier transform (STFT) front end and one evaluation harness. (ii) We build a deep sparse coding network that predicts a soft mask by unrolling a non-negative proximal-gradient iteration and learning its dictionaries end-to-end, and show it is the strongest non-oracle separator while being the fastest learned model. (iii) We pair each method with its recovery condition and report a negative result on dictionary coherence. Code, weights, and audio are released.

Related work. NMF for audio dates to Smaragdis & Brown (2003); supervised and convolutive variants (Smaragdis, 2007; Févotte et al., 2009) pre-train one dictionary per source. Robust-PCA singing-voice separation (Huang et al., 2012) and REPET (Rafii & Pardo, 2013) exploit the repetitive, low-rank structure of accompaniment. K-SVD (Aharon et al., 2006) learns overcomplete sparsifying dictionaries and SRC (Wright et al., 2009) classifies by reconstruction residual. On the deep side, Spleeter, Open-Unmix, and Demucs train end-to-end masks (Stöter et al., 2019; Défossez et al., 2021; Rouard et al., 2023), while a complementary line reads such networks as unrolled sparse coding: LISTA unrolls ISTA (Gregor & LeCun, 2010), CNNs admit a convolutional-sparse-coding reading (Papyan et al., 2017), and supervised deep sparse coding networks have been studied for image classification (Sun et al., 2019). Our learned separator instantiates this unrolling idea for music spectrogram masking.

2. Background: Spectrograms, Stems, and Masking

The short-time Fourier transform slides a window across $y(t)$ and takes a DFT in each window, producing a complex matrix $\boldsymbol{Y} \in \mathbb{C}^{F\times T}$ with F frequency bins and T frames. Its modulus $\boldsymbol{M} = |\boldsymbol{Y}|$ is a nonnegative image (brightness is energy) and its argument $\angle\boldsymbol{Y}$ is the phase; the transform is invertible. A finished song is a sum of source stems, and because the STFT is linear the mixture spectrogram is approximately the sum of the source spectrograms. Rather than predict a source spectrogram outright, we predict a mask $\boldsymbol{M}_c \in [0,1]^{F\times T}$ that scales each cell and resynthesize with the mixture phase. The minimum-mean-square-error mask under locally uncorrelated Gaussian sources is the ideal ratio mask $\boldsymbol{M}_c^{\mathrm{IRM}} = |\boldsymbol{Y}_c|^2/\sum_k |\boldsymbol{Y}_k|^2$; it needs the true sources and so serves only as an oracle ceiling. Two empirical regularities make the masks recoverable (Figure 2): a voiced frame places energy on a sparse set of harmonics, and sustained accompaniment reuses a few spectral templates and is therefore low-rank, the two structures compressed sensing is built to exploit.

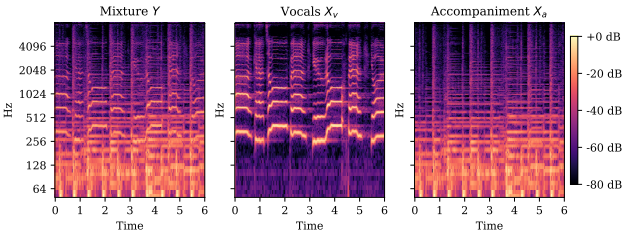

Figure 2. Log-magnitude STFT of a test clip: the mixture, the true vocals (a sparse harmonic scaffold), and the true accompaniment (denser and lower-rank). Our methods reverse the sum.

3. Preliminaries

Write $\|\boldsymbol{x}\|_0 = |\mathrm{supp}(\boldsymbol{x})|$. Given $\boldsymbol{A} \in \mathbb{R}^{m\times N}$ with $m \ll N$ and $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ for an s-sparse $\boldsymbol{x}$, the combinatorial program $\min\|\boldsymbol{z}\|_0$ s.t. $\boldsymbol{A}\boldsymbol{z} = \boldsymbol{y}$ is NP-hard; its ℓ_1 relaxation is a linear program. Three sufficient conditions certify exactness: $\mathrm{spark}(\boldsymbol{A}) > 2s$ (uniqueness); the restricted isometry property $(1-\delta_s)\|\boldsymbol{x}\|_2^2 \le \|\boldsymbol{A}\boldsymbol{x}\|_2^2 \le (1+\delta_s)\|\boldsymbol{x}\|_2^2$ with $\delta_{2s} < \sqrt{2}-1$ (Candès & Tao, 2005); and the verifiable coherence bound

$$s < \tfrac{1}{2}\left(1 + \tfrac{1}{\mu(\boldsymbol{A})}\right), \qquad \mu(\boldsymbol{A}) = \max_{i\neq j} |\langle \boldsymbol{A}_i, \boldsymbol{A}_j\rangle|, \quad (1)$$

under which orthogonal matching pursuit (OMP) and ℓ_1 minimization recover every s-sparse signal (Tropp & Gilbert, 2007; Donoho, 2006). For low rank, the nuclear norm $\|\boldsymbol{X}\|_* = \sum_i \sigma_i(\boldsymbol{X})$ is the convex surrogate, and Robust PCA splits $\boldsymbol{M} = \boldsymbol{X} + \boldsymbol{E}$ into low-rank $\boldsymbol{X}$ and sparse $\boldsymbol{E}$ by principal component pursuit (Candès et al., 2011).

4. Method

Front end. For a mono mixture $y(t)$, the STFT is $\boldsymbol{Y} \in \mathbb{C}^{F\times T}$ with magnitude $\boldsymbol{M} = |\boldsymbol{Y}|$ and phase $\angle\boldsymbol{Y}$. Every separator emits a soft mask $\boldsymbol{M}_c \in [0,1]^{F\times T}$ and resynthesizes $\hat{y}_c = \mathrm{STFT}^{-1}(\boldsymbol{M}_c \odot \boldsymbol{Y})$ with the mixture phase; given nonnegative magnitude estimates $\widehat{\boldsymbol{M}}_c$ we use Wiener masks $\boldsymbol{M}_c = \widehat{\boldsymbol{M}}_c^2/\sum_k \widehat{\boldsymbol{M}}_k^2$. The methods differ only in how they estimate $\widehat{\boldsymbol{M}}_c$ (Table 1).

(A) Robust PCA. We solve $\min_{\boldsymbol{X},\boldsymbol{E}} \|\boldsymbol{X}\|_* + \lambda\|\boldsymbol{E}\|_1$ s.t. $\boldsymbol{X} + \boldsymbol{E} = \boldsymbol{M}$, $\lambda = 1/\sqrt{\max(F,T)}$, by inexact ALM (singular-value thresholding for the nuclear prox, soft-thresholding for ℓ_1). The low-rank $\boldsymbol{X}$ is the accompaniment and the sparse $\boldsymbol{E}$ the vocal proxy, the same

Table 1. The four separators as one family: a structural prior, its penalty, and a learning mode.

Method	Prior	Penalty	Learning
RPCA	low-rank+sparse	nuclear+ℓ_1	none
NMF	nonneg. parts	KL div.	unsup.
K-SVD	sparse code	$\ell_0 \le s$	unsup.
SCN	unrolled sparse	elastic-net	end-to-end

intuition that pulls a moving foreground out of a static background in surveillance video. Training-free.

(B) Supervised NMF. We pre-train one nonnegative dictionary per source, $\boldsymbol{M}_c \approx \boldsymbol{D}_c \boldsymbol{H}_c$, by Lee-Seung KL multiplicative updates (Lee & Seung, 2000), which are a majorization-minimization scheme that cannot increase the loss. At test time we fix $\boldsymbol{D} = [\boldsymbol{D}_v \mid \boldsymbol{D}_a]$ and solve for the activations, then split $\widehat{\boldsymbol{M}}_v = \boldsymbol{D}_v \boldsymbol{H}_v$ and $\widehat{\boldsymbol{M}}_a = \boldsymbol{D}_a \boldsymbol{H}_a$.

(C) K-SVD + SRC. We learn an overcomplete dictionary per source by K-SVD (Aharon et al., 2006) (alternating OMP coding with rank-one atom updates that are Eckart-Young optimal), sparse-code each mixture frame on $[\boldsymbol{D}_v \mid \boldsymbol{D}_a]$, and route energy by the SRC residual rule $c^\star = \arg\min_c \|\boldsymbol{y} - \boldsymbol{D}_c \boldsymbol{\alpha}_c\|_2$ (Wright et al., 2009). Coherence Equation (1) governs when this per-frame code is unique.

(D) Deep sparse coding network. Our learned separator predicts the mask by unrolling a non-negative sparse solver and learning its dictionaries end-to-end. Each composite module is an expand-then-reduce hourglass (Figure 3): a fat dictionary lifts the frame to an over-complete code, a tall dictionary compresses it, each computed as

$$\boldsymbol{\alpha}^{(\ell)} = \underset{\boldsymbol{\alpha} \ge 0}{\arg\min} \tfrac{1}{2}\|\boldsymbol{u} - \boldsymbol{D}^{(\ell)}\boldsymbol{\alpha}\|_2^2 + \lambda_1 \|\boldsymbol{\alpha}\|_1 + \tfrac{\lambda_2}{2}\|\boldsymbol{\alpha}\|_2^2, \quad (2)$$

solved by K steps of non-negative FISTA, each a matrix multiply, gradient step, and clamp $\boldsymbol{\alpha}_k = [\boldsymbol{z}_k - \frac{1}{L}\boldsymbol{D}^\top(\boldsymbol{D}\boldsymbol{z}_k - \boldsymbol{u}) - \frac{\lambda_1}{L}]_+$ with Nesterov momentum. All operations are differentiable, so the dictionaries $\boldsymbol{D}^{(\ell)}$ and shrinkages λ_1, λ_2 train by backpropagation against the ideal ratio mask with an ℓ_1 penalty on the codes. This is the LISTA construction (Gregor & LeCun, 2010) specialized to non-negative codes and a masking output; the clamp is a proximal step, and run to convergence the iteration attains the optimal $O(1/k^2)$ rate (Beck & Teboulle, 2009). We stack four modules with a ± 2-frame temporal context window (1.44M parameters). A simultaneous-OMP variant gives a stereo joint-sparsity extension (Tropp et al., 2006).

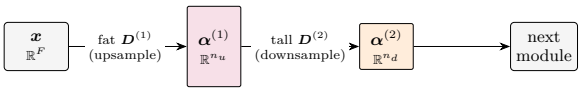

Figure 3. One composite module: a fat dictionary lifts the frame to an over-complete sparse code, a tall dictionary compresses it. The expand-then-reduce shape is the autoencoder hourglass with a sparse bottleneck.

Table 2. Median SI-SDR (dB) on 25 MUSDB18 test tracks with mean inference time per 6 s clip. Best non-oracle in bold.

Method	Vocals	Accomp.	Time (s)
Mixture (no split)	−3.7	4.0	0.00
Random mask	−3.9	2.9	0.01
HPSS (median)	−11.9	−1.4	0.25
REPET-SIM	−0.7	2.9	0.08
RPCA	−3.1	0.2	1.14
NMF (supervised)	0.5	5.1	0.46
K-SVD + SRC	0.8	7.1	0.28
SCN (ours)	**2.0**	**8.3**	**0.12**
Oracle IRM / IBM	10.2	15.8	—

5. Experiments

Setup. We use the MUSDB18 benchmark (Rafii et al., 2017), separating vocals from accompaniment. Dictionaries and the deep model are trained on the train split; all numbers are medians over 25 held-out test tracks at 22.05 kHz, 2048-point STFT, hop 512. We report scale-invariant SDR (Le Roux et al., 2019) and compare against two training-free baselines, median-filtering HPSS and REPET-SIM (FitzGerald, 2010; Rafii & Pardo, 2013); two trivial floors (returning the mixture; a random mask); and the ideal ratio/binary mask oracles. NMF uses K=64 KL atoms, K-SVD K=96, s=10, and the deep coder four modules, 1.44M parameters, trained on a laptop GPU.

Main result. Table 2 and Figure 4 show a clean ordering. The floors calibrate the scale: returning the mixture already scores +4.0 dB on accompaniment because the band dominates the energy, so accompaniment numbers read against this floor. The training-free baselines are weak (HPSS falls below even the mixture floor). The supervised dictionary methods improve sharply, and the deep sparse coding network is the best non-oracle separator on both sources (2.0 dB vocals, 8.3 dB accompaniment), beating the best classical method by $\approx$ 1.2 dB on each, while being the fastest learned model since its forward pass is a fixed number of matrix multiplies rather than an iterative solve. Soft (ratio) masks beat hard binary masks by $\approx$ 1 dB for every method. All methods stay below the

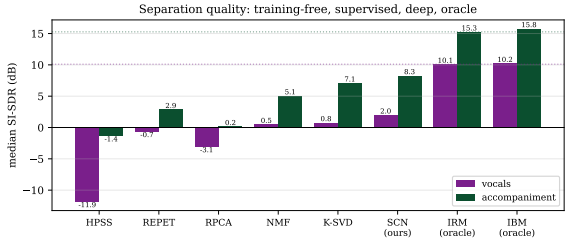

Figure 4. Median SI-SDR by method; dotted lines mark the oracle ceilings. Supervision and unrolled learning each add a clear increment over the training-free baselines.

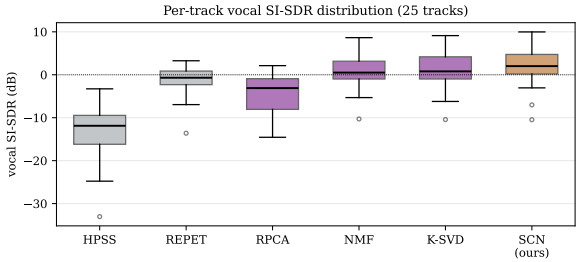

Figure 5. Per-track vocal SI-SDR over the 25 test tracks. The deep model shifts the full distribution rather than only improving easy tracks.

oracle ceilings, as any mixture-phase magnitude mask must.

Per-track spread, masks, stereo, cost. Medians hide variance, so Figure 5 plots the per-track vocal SI-SDR distribution: the deep model shifts the whole distribution rightward, not only the easy tracks, though every method has a long lower tail on dense, vocal-quiet mixtures. The masking strategy matters independently of the model: soft (ratio) masks beat hard binary masks by $\approx$ 1 dB for every method (2.1 vs 1.0 dB for the deep coder), because a continuous gain keeps the partial-overlap cells a binary decision discards. The stereo simultaneous-OMP separator (Section 4) reaches 0.9/7.3 dB on 10 stereo tracks, matching the mono dictionary it is built on while keeping the two channels consistent. Table 3 contrasts model size and per-clip cost: the deep coder is the smallest learned model we train and the fastest at test time, an order of magnitude quicker than the iterative RPCA solve, because its forward pass is a fixed number of matrix multiplies.

Ablations. Table 4 sweeps the central knobs. Supervised NMF is best at the smallest dictionary (K=16): with limited data a small dictionary is a stronger prior, a regularization effect. K-SVD improves with the sparsity budget to s=10 then saturates. RPCA improves

Table 3. Model size and per-clip cost. Classical methods store fixed dictionaries; SCN learns dictionaries and shrinkages end-to-end.

Method	Trains?	Params	Test (s)
RPCA	no	—	1.14
NMF	dict. only	0.13M	0.46
K-SVD	dict. only	0.20M	0.28
SCN (ours)	end-to-end	1.44M	0.12

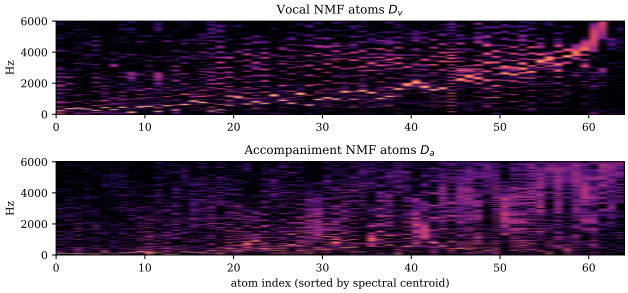

Figure 6. Supervised NMF dictionaries, atoms sorted by spectral centroid: vocal atoms (top) ride the formant region, accompaniment atoms (bottom) spread lower.

as the penalty λ grows, pushing more energy into the low-rank band and leaving a cleaner sparse vocal, so the canonical λ is conservative for this task. The learned dictionaries themselves are musically sensible: Figure 6 shows the NMF vocal atoms concentrating harmonic ridges in the formant region while the accompaniment atoms spread lower and wider, the parts-based structure NMF is designed to recover. The Robust PCA solver is well-behaved too: Figure 7 shows its residual reaching 10^{-7} in 38 iterations and the mixture's singular spectrum collapsing after a handful of components, the empirical evidence for the low-rank accompaniment prior.

Coherence: a negative result. The recovery condition Equation (1) needs low coherence, but a dictionary trained for reconstruction is coherent. Figure 8 shows the learned K-SVD vocal dictionary has $\mu \approx 0.85$, so Equation (1) certifies exact recovery only for $s < \frac{1}{2}(1 + 1/\mu) \approx 1.1$, far below the s=10 at which the method works. The worst-case guarantee is vacuous here, yet separation succeeds: coherence bounds the worst sparse vector, while typical music frames are far from that adversarial case. We report this to set honest expectations for how much classical theory transfers to learned audio dictionaries.

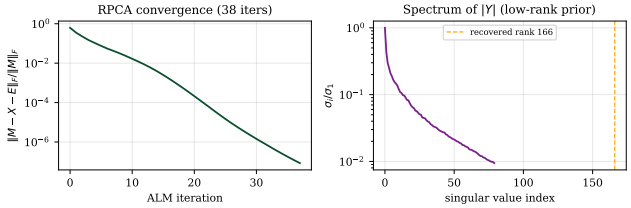

Figure 7. Left: relative residual of the principal-component-pursuit iteration. Right: normalized singular spectrum of the mixture magnitude, justifying the low-rank accompaniment prior.

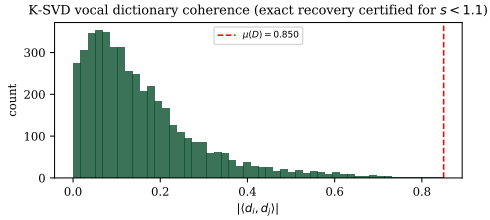

Figure 8. Off-diagonal coherence of the learned K-SVD vocal dictionary. The worst-case bound certifies recovery only for $s \lesssim 1$, but separation at s=10 works, evidence of an average-case-versus-worst-case gap.

Table 4. Ablations (median vocal SI-SDR, dB, 12 tracks).

NMF atoms K	16	32	64	128
Vocal SI-SDR	2.2	1.9	1.8	1.5
K-SVD sparsity s	3	5	10	15
Vocal SI-SDR	1.4	2.3	**2.7**	2.7
RPCA λ scale	0.5	1.0	1.5	2.0
Vocal SI-SDR	−3.9	−3.6	−3.2	**−2.3**

6. Discussion and Conclusion

The numbers are deliberately scoped: short clips, modest dictionaries, and a 1.4M parameter network trained on a laptop. Large end-to-end systems reach higher vocal SI-SDR with orders of magnitude more parameters and data (Rouard et al., 2023; Mitsufuji et al., 2022); the value here is the unifying view and the controlled comparison. On one front end, increasing structure-learning, from a training-free split to supervised factorization to a learned unrolled solver, produces a monotone improvement, and the unrolled solver is both the most accurate and the cheapest to run. Two limitations are intrinsic: magnitude masking with mixture phase caps every method at the IRM oracle, and the deep model uses only a short temporal context. Code, trained weights, the evaluation harness, and curated audio are available at https://github.com/takakhoo/ENGS109_Final_Project.

References

Aharon, M., Elad, M., and Bruckstein, A. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. IEEE Trans. Signal Processing, 54(11):4311–4322, 2006.

Beck, A. and Teboulle, M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM Journal on Imaging Sciences, 2(1):183–202, 2009.

Candès, E. J. and Tao, T. Decoding by linear programming. IEEE Trans. Information Theory, 51 (12):4203–4215, 2005.

Candès, E. J., Li, X., Ma, Y., and Wright, J. Robust principal component analysis? Journal of the ACM, 58(3):1–37, 2011.

Cano, E., FitzGerald, D., Liutkus, A., Plumbley, M. D., and Stöter, F.-R. Musical source separation: An introduction. IEEE Signal Processing Magazine, 36(1):31–40, 2019.

Défossez, A., Usunier, N., Bottou, L., and Bach, F. Music source separation in the waveform domain. arXiv:1911.13254, 2021.

Donoho, D. L. Compressed sensing. IEEE Trans. Information Theory, 52(4):1289–1306, 2006.

Févotte, C., Bertin, N., and Durrieu, J.-L. Nonnegative matrix factorization with the Itakura-Saito divergence. Neural Computation, 21(3):793–830, 2009.

FitzGerald, D. Harmonic/percussive separation using median filtering. In Int. Conf. on Digital Audio Effects (DAFx), 2010.

Gregor, K. and LeCun, Y. Learning fast approximations of sparse coding. In Int. Conf. on Machine Learning (ICML), pp. 399–406, 2010.

Huang, P.-S., Chen, S. D., Smaragdis, P., and Hasegawa-Johnson, M. Singing-voice separation from monaural recordings using robust principal component analysis. In IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP), pp. 57–60, 2012.

Le Roux, J., Wisdom, S., Erdogan, H., and Hershey, J. R. SDR – half-baked or well done? In IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP), pp. 626–630, 2019.

Lee, D. D. and Seung, H. S. Learning the parts of objects by non-negative matrix factorization. Nature, 401(6755):788–791, 1999.

Lee, D. D. and Seung, H. S. Algorithms for non-negative matrix factorization. In Advances in Neural Information Processing Systems (NeurIPS), pp. 556–562, 2000.

Mitsufuji, Y., Fabbro, G., Uhlich, S., Stöter, F.-R., Défossez, A., Kim, M., and Choi, W. Music demixing challenge 2021. Frontiers in Signal Processing, 1:808395, 2022.

Monga, V., Li, Y., and Eldar, Y. C. Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing. IEEE Signal Processing Magazine, 38(2):18–44, 2021.

Papyan, V., Romano, Y., and Elad, M. Convolutional neural networks analyzed via convolutional sparse coding. Journal of Machine Learning Research, 18 (83):1–52, 2017.

Rafii, Z. and Pardo, B. REPET: A simple method for music/voice separation. IEEE Trans. Audio, Speech, Language Process., 21(1):73–84, 2013.

Rafii, Z., Liutkus, A., Stöter, F.-R., Mimilakis, S. I., and Bittner, R. The MUSDB18 corpus for music separation, 2017.

Rouard, S., Massa, F., and Défossez, A. Hybrid transformers for music source separation. In IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP), 2023.

Smaragdis, P. Convolutive speech bases and their application to supervised speech separation. IEEE Trans. Audio, Speech, Language Process., 15(1):1–12, 2007.

Smaragdis, P. and Brown, J. C. Non-negative matrix factorization for polyphonic music transcription. In IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), pp. 177–180, 2003.

Stöter, F.-R., Uhlich, S., Liutkus, A., and Mitsufuji, Y. Open-unmix – a reference implementation for music source separation. Journal of Open Source Software, 4(41):1667, 2019.

Sun, X., Nasrabadi, N. M., and Tran, T. D. Supervised deep sparse coding networks for image classification. IEEE Trans. Image Processing, 29:405–418, 2019.

Tropp, J. A. and Gilbert, A. C. Signal recovery from random measurements via orthogonal matching pursuit. IEEE Trans. Information Theory, 53(12):4655–4666, 2007.

Tropp, J. A., Gilbert, A. C., and Strauss, M. J. Algorithms for simultaneous sparse approximation. part i: Greedy pursuit. Signal Processing, 86(3):572–588, 2006.

Wright, J., Yang, A. Y., Ganesh, A., Sastry, S. S., and Ma, Y. Robust face recognition via sparse representation. IEEE Trans. Pattern Analysis and Machine Intelligence, 31(2):210–227, 2009.