Robust Deconvolution with Parseval Filterbanks

Rossen Nenov*, Vincent Lostanlen†, and Peter Balazs*‡
*Acoustics Research Institute, Austrian Academy of Sciences, Vienna, Austria
†Nantes Université, École Centrale Nantes, LS2N, CNRS, UMR 6004, F-44000 Nantes, France
‡ Interdisciplinary Transformation University Austria (IT:U), Linz, Austria

rossen.nenov@oeaw.ac.at

***Abstract*—This article introduces two contributions: Multi-band Robust Deconvolution (Multi-RDCP), a regularization approach for deconvolution in the presence of noise; and Subband-Normalized Adaptive Kernel Evaluation (SNAKE), a first-order iterative algorithm designed to efficiently solve the resulting optimization problem. Multi-RDCP resembles Group LASSO in that it promotes sparsity across the subband spectrum of the solution. We prove that SNAKE enjoys fast convergence rates and numerical simulations illustrate the efficiency of SNAKE for deconvolving noisy oscillatory signals.**

***Index Terms*—Deconvolution, digital signal processing, gradient methods, signal reconstruction, sparse approximation.**

I. Introduction

Deconvolution is a well-known inverse problem with numerous applications in signal processing [1], [2]. The (circular) convolution between two finite sequences $\boldsymbol{x}$ and $\boldsymbol{w}$ in $\mathbb{R}^N$ is:

$$(\boldsymbol{w} * \boldsymbol{x})[n] = \sum_{m=0}^{N-1} \boldsymbol{w}[m]\boldsymbol{x}\big[(n-m) \bmod N\big]. \qquad (1)$$

Reciprocally, (non-blind) deconvolution operators assume the knowledge of two finite sequences $\boldsymbol{x}, \boldsymbol{y}$, and seek $\boldsymbol{w}$ such that $\boldsymbol{w} * \boldsymbol{x} \approx \boldsymbol{y}$; more precisely, such that a certain cost function F is minimized. One such function is mean square error (MSE):

$$f : \boldsymbol{w} \in \mathbb{R}^N \longmapsto \tfrac{1}{2}\big\|\boldsymbol{w} * \boldsymbol{x} - \boldsymbol{y}\big\|^2. \qquad (2)$$

If one simply chooses to minimize Equation (2), deconvolution can be solved directly via the fast Fourier transform (FFT) [3] by spectral division. Denoting the discrete Fourier transform of $\boldsymbol{x}$ by $\mathcal{F}(\boldsymbol{x}) = \hat{\boldsymbol{x}}$, the convolution theorem [4] yields, for every frequency k: $\mathcal{F}(\boldsymbol{w} * \boldsymbol{x})[k] = \hat{\boldsymbol{w}}[k]\hat{\boldsymbol{x}}[k]$. Let $\boldsymbol{w}^* = \mathcal{F}^{-1}(\hat{\boldsymbol{w}}^*)$ where, for every k, $\hat{\boldsymbol{w}}^*[k] = \hat{\boldsymbol{y}}[k]/\hat{\boldsymbol{x}}[k]$ if $\hat{\boldsymbol{x}}[k] \neq 0$ and $\hat{\boldsymbol{w}}^*[k] = 0$ otherwise. One may verify that $\boldsymbol{w}^*$ minimizes the MSE. Indeed, f is convex and its gradient is zero at $\boldsymbol{w}^*$ [4].

A different approach is needed for regularized deconvolution problems, whose cost function writes as $F = f + r$. Here, f is a data fidelity term–like MSE–depending on $\boldsymbol{x}$ and $\boldsymbol{y}$, while r penalizes a certain notion of "complexity" in the solution. The practical interest behind these problems lies in the need to reliably estimate the impulse response of the linear time–invariant (LTI) system which maps $\boldsymbol{x}$ to $\boldsymbol{y}$ in the presence of highly noisy observations. The complexity of the solution may arguably be quantified in terms of spectral spread, i.e., the notion that the entries $|\hat{\boldsymbol{w}}[k]|^2$ are non-negligible for a wide range of frequency values k. Narrowband LTI systems are ubiquitous in applied sciences: formants in phonetics [5] and Helmholtz resonators in musical acoustics [6] are some well-known examples.

In this article, we introduce the *multiband robust deconvolution problem* (Multi-RDCP). The key idea is to employ a predefined filterbank $\boldsymbol{\Psi}$ to decompose $\boldsymbol{x}$ and $\boldsymbol{y}$ into $J > 1$ subbands. This decomposition enables the formulation of individual subproblems and the regularization of the solution $\boldsymbol{w}$ through a non-squared ℓ^2-norm penalty applied to each subband component, promoting sparsity across subbands, akin to Group LASSO [7]. Since Multi-RDCP lacks closed-form solutions and full differentiability, an iterative method such as proximal gradient descent (PGD) [8] is required. Yet, we will see that general PGD is impractically slow if $\hat{\boldsymbol{x}}$ exhibits large variations in magnitude across regions of the Fourier spectrum. In order to overcome this, we present a new algorithm: *subband-normalized adaptive kernel evaluation (SNAKE)*. SNAKE applies PGD over each subband independently, with a step size that is adapted to the power spectral density of $\boldsymbol{x}$ within that band.

Our main theoretical result (Theorem III.6) is that SNAKE enjoys a global linear convergence rate. We corroborate this result through numerical experiments, demonstrating significant convergence speedups compared to classical PGD, along with an improved robustness to noise.

Note. We write $[N]$ as a shorthand for $\{0, \ldots, N-1\}$. We denote the time reversal of a finite sequence $\boldsymbol{x} \in \mathbb{R}^N$ by $\tilde{\boldsymbol{x}}$.

II. Proximal gradient-based deconvolution

Definition II.1. Given $\lambda > 0$ and $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^N$, the *robust deconvolution problem (RDCP)* of $\boldsymbol{y}$ from $\boldsymbol{x}$ minimizes

$$F : \boldsymbol{w} \in \mathbb{R}^N \longmapsto \tfrac{1}{2}\left\|\boldsymbol{w} * \boldsymbol{x} - \boldsymbol{y}\right\|_2^2 + \lambda\|\boldsymbol{w}\|_2, \qquad (3)$$

i.e., $F = f + r$ with f is the MSE and r the ℓ^2-norm times λ.

Note. We square the ℓ^2 norm in f but not in r. This is an important design choice, which we will discuss in Section III.

A common choice for minimizing the sum of a smooth and a non-smooth convex function is proximal gradient descent.

Definition II.2 ([9])**.** The *proximal operator* of a convex, lower semi-continuous function $g : \mathbb{R}^N \to \mathbb{R}$ is defined as

$$\mathrm{Prox}_{\alpha g} : \boldsymbol{w} \in \mathbb{R}^N \longmapsto \operatorname*{arg\,min}_{\boldsymbol{v} \in \mathbb{R}^N} \left\{ g(\boldsymbol{v}) + \tfrac{1}{2\alpha}\|\boldsymbol{v} - \boldsymbol{w}\|_2^2 \right\}. \qquad (4)$$

Corresponding author: Rossen Nenov. The work of R. Nenov and P. Balazs is supported by the Austrian Science Fund (FWF) projects LoFT [10.55776/P34624], NoMASP [10.55776/P34922] and Voice Prints [10.55776/P36446] and the WWTF project EleCom [10.47379/LS23024]. The work of V. Lostanlen is supported by ANR project MuReNN (ANR-23-CE23-0007-01). To reproduce our numerical simulations, please visit: https://github.com/RosseN16/SNAKE

Lemma II.3 ([10], Section 6.5.1)**.** *Given $\alpha > 0$ and $\lambda > 0$, the proximal operator of $r : \boldsymbol{w} \in \mathbb{R}^N \mapsto \lambda \|\boldsymbol{w}\|_2$ is given by*

$$\mathrm{Prox}_{\alpha r} : \boldsymbol{w} \longmapsto \max\big(0, 1 - \alpha\lambda / \|\boldsymbol{w}\|_2\big)\boldsymbol{w}. \tag{5}$$

Definition II.4 ([8])**.** Given $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^N$ and RDCP (Def. II.1), an initial guess $\boldsymbol{w}^{(0)} \in \mathbb{R}^N$ and a step size $\alpha > 0$, *proximal gradient descent (PGD)* computes the sequence $\{\boldsymbol{w}^{(t)}\}_{t\in\mathbb{N}}$ recursively:

$$\boldsymbol{w}^{(t+1)} = \mathrm{Prox}_{\alpha r}\left(\boldsymbol{w}^{(t)} - \alpha \boldsymbol{\nabla} f(\boldsymbol{w}^{(t)})\right). \tag{6}$$

PGD theory has first been developed under the assumption of strong convexity. More recently, [11] have proved the linear convergence of PGD under a weaker assumption than strong convexity, called "proximal Polyak-Łojasiewicz" (Prox-PŁ) in reference to [12] and [13]. We proceed to establish the same result under an even weaker assumption than Prox-PŁ.

First we state some results on the MSE f in RDCP.

Definition II.5 ([14])**.** For a $\boldsymbol{x} \in \mathbb{R}^N$ the associated *circulant matrix* $\mathbf{C}_{\boldsymbol{x}} \in \mathbb{R}^{N\times N}$, such that $\boldsymbol{w} * \boldsymbol{x} = \mathbf{C}_{\boldsymbol{x}}\boldsymbol{w}$, is given by $\mathbf{C}_{\boldsymbol{x}}[n, n'] = \boldsymbol{x}[(n - n') \mod N]$ for $n, n' \in [N]$.

Definition II.6. Given $\boldsymbol{x} \in \mathbb{R}^N$, we denote its smallest nonzero and largest power spectrum entries by

$$S^+_{\min}(\boldsymbol{x}) = \min_{k\in[N]} \left\{|\hat{\boldsymbol{x}}|^2[k] \,\middle|\, |\hat{\boldsymbol{x}}|^2[k] > 0\right\}, \tag{7}$$

$$S_{\max}(\boldsymbol{x}) = \max_{k\in[N]} \left\{|\hat{\boldsymbol{x}}|^2[k]\right\}, \tag{8}$$

where $\hat{\boldsymbol{x}}$ is the discrete Fourier transform of $\boldsymbol{x}$.

Lemma II.7. *Given $\boldsymbol{w}, \boldsymbol{v} \in \mathbb{R}^N$, Equation (2) satisfies: $f(\boldsymbol{v}) = f(\boldsymbol{w}) + \langle \boldsymbol{\nabla} f(\boldsymbol{w}), \boldsymbol{v} - \boldsymbol{w}\rangle + \frac{1}{2}\|\boldsymbol{v} * \boldsymbol{x} - \boldsymbol{w} * \boldsymbol{x}\|_2^2$.*

Proof. We rewrite f in terms of $\mathbf{C}_{\boldsymbol{x}}$ and expand the norm to obtain $f(\boldsymbol{v}) = \frac{1}{2}(\|\mathbf{C}_{\boldsymbol{x}}\boldsymbol{w} - \boldsymbol{y}\|_2^2 + 2\langle \mathbf{C}_{\boldsymbol{x}}\boldsymbol{w} - \boldsymbol{y}, \mathbf{C}_{\boldsymbol{x}}\boldsymbol{v} - \mathbf{C}_{\boldsymbol{x}}\boldsymbol{w}\rangle + \|\mathbf{C}_{\boldsymbol{x}}\boldsymbol{v} - \mathbf{C}_{\boldsymbol{x}}\boldsymbol{w}\|_2^2)$. Since the gradient is given by $\boldsymbol{\nabla} f(\boldsymbol{w}) = \mathbf{C}_{\boldsymbol{x}}^\top(\mathbf{C}_{\boldsymbol{x}}\boldsymbol{w} - \boldsymbol{y})$, we conclude with shifting the transpose in $\langle \boldsymbol{\nabla} f(\boldsymbol{w}), \boldsymbol{w} - \boldsymbol{v}\rangle = \langle \mathbf{C}_{\boldsymbol{x}}\boldsymbol{w} - \boldsymbol{y}, \mathbf{C}_{\boldsymbol{x}}\boldsymbol{w} - \mathbf{C}_{\boldsymbol{x}}\boldsymbol{v}\rangle$. ☐

Lemma II.8. *For $\boldsymbol{v} \in \mathbb{R}^N$, $\|\boldsymbol{v} * \boldsymbol{x}\|_2^2 \le S_{\max}(\boldsymbol{x})\,\|\boldsymbol{v}\|_2^2$. For all $\boldsymbol{w} \in \ker(\mathbf{C}_{\boldsymbol{x}})^\perp$, $S^+_{\min}(\boldsymbol{x})\,\|\boldsymbol{w}\|_2^2 \le \|\boldsymbol{w} * \boldsymbol{x}\|_2^2$.*

Proof. The matrix $\mathbf{C}_{\boldsymbol{x}}$ is diagonalized by the discrete Fourier transform and its singular values are the entries of $\hat{\boldsymbol{x}}$ [15]. ☐

We now introduce a measure of the distance of $\boldsymbol{w}$ to the set of optimal solutions in terms of function values and we establish its connection to the iterates of PGD for RDCP.

Definition II.9. The *proximal gradient optimality gap* of RDCP (Def. II.1), for $\boldsymbol{w} \in \mathbb{R}^N$ and $\gamma > 0$, is given by:

$$D_r(\boldsymbol{w}, \gamma) := \tag{9}$$
$$-\gamma \min_{\boldsymbol{v}\in\mathcal{V}} \left\{\langle \boldsymbol{\nabla} f(\boldsymbol{w}), \boldsymbol{v} - \boldsymbol{w}\rangle + \tfrac{\gamma}{2}\|\boldsymbol{v} - \boldsymbol{w}\|_2^2 + r(\boldsymbol{v}) - r(\boldsymbol{w})\right\},$$

where $\mathcal{V} = \mathrm{ran}(\mathbf{C}_{\boldsymbol{x}}^\top)$ is the row space of the matrix $\mathbf{C}_{\boldsymbol{x}}$.

Note. This definition is similar to [11, Equation 13] except that we restrict the minimum to $\mathcal{V}$, i.e., a linear subspace of $\mathbb{R}^N$. For $\boldsymbol{v} \in \mathcal{V}$, we have $\|\boldsymbol{v} * \boldsymbol{x} - \boldsymbol{w} * \boldsymbol{x}\|_2^2 > 0$ unless $\boldsymbol{v} = \boldsymbol{w}$. Indeed, $\mathcal{V}$ is orthogonal to $\ker(\mathbf{C}_{\boldsymbol{x}})$, along which f is constant.

Lemma II.10. *The proximal gradient optimality gap of RDCP (Def. II.1) is monotonically nondecreasing with respect to its second argument: $\forall \boldsymbol{w} \in \mathbb{R}^N$, $\forall \gamma, \gamma' > 0$:*

$$\big(\gamma < \gamma'\big) \implies \big(D_r(\boldsymbol{w}, \gamma) \le D_r(\boldsymbol{w}, \gamma')\big). \tag{10}$$

Proof. This can be proven analogously to [11, Appendix E, Lemma 1] by ensuring that the same arguments apply if the minimum is taken over $\mathcal{V} = \mathrm{ran}(\mathbf{C}_{\boldsymbol{x}}^\top)$ instead of over $\mathbb{R}^N$. ☐

Lemma II.11. *Given $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^N$ and RDCP (Def. II.1), an initial guess $\boldsymbol{w}^{(0)} \in \mathrm{ran}(\mathbf{C}_{\boldsymbol{x}}^\top)$ and a step size $\alpha = \frac{1}{L}$. Then the iterates of PGD satisfy $\boldsymbol{w}^{(t)} \in \mathrm{ran}(\mathbf{C}_{\boldsymbol{x}}^\top)$ for $t \in \mathbb{N}$. Furthermore, the following holds:*

$$\begin{aligned}\langle \nabla f(\boldsymbol{w}^{(t)}), &\boldsymbol{w}^{(t+1)} - \boldsymbol{w}^{(t)}\rangle + \tfrac{L}{2}\|\boldsymbol{w}^{(t+1)} - \boldsymbol{w}^{(t)}\|_2^2 \\ &+ r(\boldsymbol{w}^{(t+1)}) - r(\boldsymbol{w}^{(t)}) = -\tfrac{1}{L} D_r(\boldsymbol{w}^{(t)}, L).\end{aligned}$$

Proof. Using Lemma II.3, we verify that, for any given $\boldsymbol{w} \in \mathrm{ran}(\mathbf{C}_{\boldsymbol{x}}^\top)$, $\nabla f(\boldsymbol{w})$ and $\mathrm{prox}_{\alpha r}(\boldsymbol{w})$ are both in $\mathrm{ran}(\mathbf{C}_{\boldsymbol{x}}^\top)$. Thus, starting from $\boldsymbol{w}^{(0)} \in \mathrm{ran}(\mathbf{C}_{\boldsymbol{x}}^\top)$, we prove by recurrence that $\boldsymbol{w}^{(t)} \in \mathrm{ran}(\mathbf{C}_{\boldsymbol{x}}^\top)$ for every $t \in \mathbb{N}$. Furthermore, due to the definition of the proximal operator and $\boldsymbol{w}^{(t+1)}$, we have

$$\boldsymbol{w}^{(t+1)} = \operatorname*{arg\,min}_{\boldsymbol{v}\in\mathbb{R}^N}\big\{r(\boldsymbol{v}) + \tfrac{L}{2}\|\boldsymbol{v} - \boldsymbol{w}^{(t)} + \tfrac{1}{L}\boldsymbol{\nabla} f(\boldsymbol{w}^{(t)})\|_2^2\big\}$$
$$= \operatorname*{arg\,min}_{\boldsymbol{v}\in\mathbb{R}^N}\big\{r(\boldsymbol{v}) + \langle \boldsymbol{\nabla} f(\boldsymbol{w}^{(t)}), \boldsymbol{v} - \boldsymbol{w}^{(t)}\rangle + \tfrac{L}{2}\|\boldsymbol{v} - \boldsymbol{w}^{(t)}\|_2^2\big\}$$

with the omission of $\frac{1}{L}\|\boldsymbol{\nabla} f(\boldsymbol{w}^{(t)})\|^2$, a term which does not affect the argmin. From Definition II.9, we see that:

$$\begin{aligned}&-\tfrac{1}{L} D_r(\boldsymbol{w}^{(t)}, L) + r(\boldsymbol{w}^{(t)}) = \\ &\min_{\boldsymbol{v}\in\mathcal{V}}\Big[r(\boldsymbol{v}) + \langle \boldsymbol{\nabla} f(\boldsymbol{w}^{(t)}), \boldsymbol{v} - \boldsymbol{w}^{(t)}\rangle + \tfrac{L}{2}\|\boldsymbol{v} - \boldsymbol{w}^{(t)}\|_2^2\Big].\end{aligned}$$

The minimum is reached for $\boldsymbol{v} = \boldsymbol{w}^{(t+1)}$, which we have proven to be in $\mathcal{V}$; thus concluding the second statement. ☐

Lemma II.12. *Given $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^N$ and RDCP (Def. II.1). Then $\arg\min_{\boldsymbol{v}\in\mathbb{R}^N} F(\boldsymbol{v}) \subset \mathrm{ran}(\mathbf{C}_{\boldsymbol{x}}^\top)$.*

Proof. Since $F(\boldsymbol{v}) \to +\infty$ if $\|\boldsymbol{v}\|_2 \to +\infty$, $\arg\min_{\boldsymbol{v}\in\mathbb{R}^N} F$ is nonempty [16]. Let $\boldsymbol{v}^* \in \arg\min_{\boldsymbol{v}\in\mathbb{R}^N} F(\boldsymbol{v})$. If $\boldsymbol{v}^* = 0$, the result is trivial. Otherwise, we write the optimality condition: $0 = \nabla f(\boldsymbol{v}^*) + \partial r(\boldsymbol{v}^*) = \mathbf{C}_{\boldsymbol{x}}^\top(\mathbf{C}_{\boldsymbol{x}}\boldsymbol{v}^* - \boldsymbol{y}) + (\lambda/\|\boldsymbol{v}^*\|_2)\boldsymbol{v}^*$. ☐

Proposition II.13. Given $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^N$ and RDCP (Def. II.1), for every $\boldsymbol{w} \in \mathrm{ran}(\mathbf{C}_{\boldsymbol{x}}^\top)$:

$$D_r(\boldsymbol{w}, L) \ge \mu\big(F(\boldsymbol{w}) - F^*\big), \tag{11}$$

where $\mu = S^+_{\min}(\boldsymbol{x})$, $L = S_{\max}(\boldsymbol{x})$, and $F^* = \min_{\mathbb{R}^N} F$.

Proof. For $\boldsymbol{v}, \boldsymbol{w} \in \mathrm{ran}(\mathbf{C}_{\boldsymbol{x}}^\top) = \ker(\mathbf{C}_{\boldsymbol{x}})^\perp$, Lemma II.8 yields $\frac{1}{2}\|\boldsymbol{v} * \boldsymbol{x} - \boldsymbol{w} * \boldsymbol{x}\|_2^2 \ge \frac{\mu}{2}\|\boldsymbol{v} - \boldsymbol{w}\|_2^2$. Plugging this into Lemma II.7 yields: $f(\boldsymbol{v}) \ge f(\boldsymbol{w}) + \langle \boldsymbol{\nabla} f(\boldsymbol{w}), \boldsymbol{v} - \boldsymbol{w}\rangle + \frac{\mu}{2}\|\boldsymbol{v} - \boldsymbol{w}\|_2^2$. We add $r(\boldsymbol{v})$, take the minimum over all elements $\boldsymbol{v} \in \mathrm{ran}(\mathbf{C}_{\boldsymbol{x}}^\top)$, and recognize Definition II.9:

$$\min_{\boldsymbol{v}\in\mathrm{ran}(\mathbf{C}_{\boldsymbol{x}}^\top)} F(\boldsymbol{v}) \ge F(\boldsymbol{w}) - \frac{1}{\mu} D_r(\boldsymbol{w}, \mu).$$

On the left, we recognize F^* (Lemma II.12). On the right, we use Lemma II.10 to replace $D_r(\boldsymbol{w}, \mu)$ by $D_r(\boldsymbol{w}, L)$. ☐

Theorem II.14. *Given $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^N$ and RDCP (Def. II.1), an initial guess $\boldsymbol{w}^{(0)} \in \operatorname{ran}(\mathbf{C}_{\boldsymbol{x}}^\top)$ and a step size $\alpha = 1/S_{\max}(\boldsymbol{x})$, the iterates $\{\boldsymbol{w}^{(t)}\}_{t\in\mathbb{N}}$ of PGD satisfy:*

$$\frac{F\big(\boldsymbol{w}^{(t)}\big) - F^*}{F\big(\boldsymbol{w}^{(0)}\big) - F^*} \leq \left(1 - \frac{S^+_{\min}(\boldsymbol{x})}{S_{\max}(\boldsymbol{x})}\right)^t. \tag{12}$$

Proof. We adapt the proof of [11, Theorem 5]. Applying Lemma II.7 for $\boldsymbol{v} = \boldsymbol{w}^{(t+1)}$ and $\boldsymbol{w} = \boldsymbol{w}^{(t)}$, and Lemma II.7 for $\frac{1}{2}\|\mathbf{C}_{\boldsymbol{x}}(\boldsymbol{w}^{(t+1)} - \boldsymbol{w}^{(t)})\|_2^2 \leq \frac{L}{2}\|\boldsymbol{w}^{(t+1)} - \boldsymbol{w}^{(t)}\|_2^2$ yield:

$$\begin{aligned} f(\boldsymbol{w}^{(t+1)}) \leq f(\boldsymbol{w}^{(t)}) &+ \langle \nabla f(\boldsymbol{w}^{(t)}), \boldsymbol{w}^{(t+1)} - \boldsymbol{w}^{(t)}\rangle \\ &+ \tfrac{L}{2}\|\boldsymbol{w}^{(t+1)} - \boldsymbol{w}^{(t)}\|_2^2. \end{aligned}$$

Combining this with Lemma II.11 and Proposition II.13 yields:

$$\begin{aligned} F(\boldsymbol{w}^{(t+1)}) &= f(\boldsymbol{w}^{(t+1)}) + r(\boldsymbol{w}^{(t)}) + r(\boldsymbol{w}^{(t+1)}) - r(\boldsymbol{w}^{(t)}) \\ &\leq F(\boldsymbol{w}^{(t)}) + \langle \nabla f(\boldsymbol{w}^{(t)}), \boldsymbol{w}^{(t+1)} - \boldsymbol{w}^{(t)}\rangle \\ &\quad + \tfrac{L}{2}\|\boldsymbol{w}^{(t+1)} - \boldsymbol{w}^{(t)}\|_2^2 + r(\boldsymbol{w}^{(t+1)}) - r(\boldsymbol{w}^{(t)}) \\ &= F(\boldsymbol{w}^{(t)}) - \tfrac{1}{L} D_r(\boldsymbol{w}^{(t)}, L) \\ &\leq F(\boldsymbol{w}^{(t)}) - \tfrac{\mu}{L}\big[F(\boldsymbol{w}^{(t)}) - F^*\big]. \end{aligned}$$

This gives us $F(\boldsymbol{w}^{(t+1)}) - F^* \leq (1 - \frac{\mu}{L})\big[F(\boldsymbol{w}^{(t)}) - F^*\big]$, from which we can conclude the statement by recurrence. ☐

Theorem II.14 guarantees that PGD applied to robust deconvolution of $\boldsymbol{y}$ from $\boldsymbol{x}$ converges in terms of optimal function value with a rate of $\big(1 - S^+_{\min}(\boldsymbol{x})/S_{\max}(\boldsymbol{x})\big)^t$ or better.

III. Filterbank analysis and normalization

In many applications, the ratio between the smallest nonzero and large power spectrum coefficients of $\boldsymbol{x}$ is close to zero, hence a poor convergence rate guarantee in Equation (12). Against this issue, we use a filterbank to partition RDCP into subproblems, each focusing on a *subband*, i.e., a subset of frequency components. By assuming that the power spectral density of $\boldsymbol{x}$ has small relative variations within each subband, we obtain subband-wise power spectrum ratios which are closer to one, yielding improved convergence rate guarantees.

Definition III.1. A *Parseval filterbank* is a collection of finite sequences $\boldsymbol{\psi}_0, \ldots, \boldsymbol{\psi}_{J-1} \in \mathbb{R}^N$ such that, for every $\boldsymbol{x} \in \mathbb{R}^N$:

$$\sum_{j=0}^{J-1} \|\boldsymbol{\psi}_j * \boldsymbol{x}\|_2^2 = \|\boldsymbol{x}\|_2^2. \tag{13}$$

Theorem III.2 ([17])**.** *Given $\boldsymbol{w} \in \mathbb{R}^N$ and a Parseval filterbank $(\boldsymbol{\psi}_0, \ldots, \boldsymbol{\psi}_{J-1})$: $\boldsymbol{w} = \sum_{j=0}^{J-1} \mathbf{C}_{\boldsymbol{\psi}_j}^\top \mathbf{C}_{\boldsymbol{\psi}_j} \boldsymbol{w}$.*

Definition III.3. Let $\boldsymbol{\Psi} = (\boldsymbol{\psi}_0, \ldots, \boldsymbol{\psi}_{J-1})$ be a Parseval filterbank. Choose $(\lambda_j)_{j=0}^{J-1}$ and $(L_j)_{j=0}^{J-1}$ such that $\lambda_j \geq 0$ and $L_j > 0$ for every $j \in [J]$. Given $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^N$, for each $j \in [J]$, define the data fidelity terms

$$f_j(\boldsymbol{w}_j) = \frac{1}{2L_j}\left\|(\boldsymbol{w}_j * \boldsymbol{\psi}_j * \boldsymbol{x}) - \sqrt{L_j}(\boldsymbol{\psi}_j * \boldsymbol{y})\right\|_2^2 \tag{14}$$

and the regularization terms $r_j(\boldsymbol{w}_j) = \lambda_j\|\boldsymbol{w}_j\|_2$. The *multi-band robust deconvolution (Multi-RDCP)* cost function at $\mathbf{W} = (\boldsymbol{w}_j)_{j=0}^{J-1}$ is given by:

$$F_{\boldsymbol{\Psi}}(\mathbf{W}) = \sum_{j=0}^{J-1} f_j(\boldsymbol{w}_j) + r_j(\boldsymbol{w}_j). \tag{15}$$

Note. The double convolutions above, $(\boldsymbol{w}_j * \boldsymbol{\psi}_j * \boldsymbol{x})$, form a *hybrid filterbank* [18]; with dyadic subsampling, it would resemble a *multiresolution neural network (MuReNN)* [19].

Note. The term $\sum_{j=0}^{J-1} \lambda_j \|\boldsymbol{w}_j\|_2$ corresponds to weighted Group LASSO [7] where the groups are the subbands of $\boldsymbol{\Psi}$.

In order to synthesize the global $\boldsymbol{w}^*$ from $\mathbf{W}$, we use the formula motivated by the following proposition.

Proposition III.4. Given $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^N$, let $\mathbf{W}^* = (\boldsymbol{w}_j^*)_{j=0}^{J-1}$ be a global optimum of unregularized Multi-RDCP (Def. III.3), i.e. $\lambda_j = 0$ for $j \in [J]$. Then

$$\boldsymbol{w}^* = \sum_{j=0}^{J-1} \frac{1}{\sqrt{L_j}} \mathbf{C}_{\boldsymbol{\psi}_j}^\top \mathbf{C}_{\boldsymbol{\psi}_j} \boldsymbol{w}_j^* \tag{16}$$

is a global optimum for the deconvolution of $\boldsymbol{y}$ from $\boldsymbol{x}$.

Proof. At the global optimum, one has $\boldsymbol{\nabla} F_{\boldsymbol{\Psi}}\big(\mathbf{W}^*\big) = 0$, thus:

$$L_j \boldsymbol{\nabla} f_j(\boldsymbol{w}_j^*) = \mathbf{C}_{(\boldsymbol{\psi}_j * \boldsymbol{x})}^\top \big(\mathbf{C}_{(\boldsymbol{\psi}_j * \boldsymbol{x})} \boldsymbol{w}_j^* - \sqrt{L_j}\mathbf{C}_{\boldsymbol{\psi}_j}^\top \boldsymbol{y}\big) = 0$$

for all $j \in [J]$. The convolution is associative and commutative:

$$\mathbf{C}_{(\boldsymbol{\psi}_j * \boldsymbol{x})}^\top \mathbf{C}_{(\boldsymbol{\psi}_j * \boldsymbol{x})} = \mathbf{C}_{\boldsymbol{\psi}_j}^\top \mathbf{C}_{\boldsymbol{x}}^\top \mathbf{C}_{\boldsymbol{\psi}_j} \mathbf{C}_{\boldsymbol{x}} = \mathbf{C}_{\boldsymbol{x}}^\top \mathbf{C}_{\boldsymbol{x}} \mathbf{C}_{\boldsymbol{\psi}_j}^\top \mathbf{C}_{\boldsymbol{\psi}_j}.$$

Thus, by dividing each identity by $\sqrt{L_j}$ and summing over j:

$$\sum_{j=0}^{J-1} \mathbf{C}_{\boldsymbol{x}}^\top \mathbf{C}_{\boldsymbol{x}} \left(\frac{1}{\sqrt{L_j}} \mathbf{C}_{\boldsymbol{\psi}_j}^\top \mathbf{C}_{\boldsymbol{\psi}_j} \boldsymbol{w}_j^*\right) = \sum_{j=0}^{J-1} \mathbf{C}_{\boldsymbol{x}}^\top \mathbf{C}_{\boldsymbol{\psi}_j}^\top \mathbf{C}_{\boldsymbol{\psi}_j} \boldsymbol{y}.$$

We recognize $\mathbf{C}_{\boldsymbol{x}}^\top \mathbf{C}_{\boldsymbol{x}} \boldsymbol{w}^*$ on the left and $\mathbf{C}_{\boldsymbol{x}}^\top \boldsymbol{y}$ on the right (see Theorem III.2). Hence: $\mathbf{C}_{\boldsymbol{x}}^\top\big(\mathbf{C}_{\boldsymbol{x}} \boldsymbol{w} - \boldsymbol{y}\big) = 0$, i.e., $\boldsymbol{\nabla} f(\boldsymbol{w}^*) = 0$ (see Equation 2). We conclude by convexity of the MSE. ☐

As the objective of Multi-RDCP is separable in $(\boldsymbol{w}_j)_{j=0}^{J-1}$, it can be solved with PGD for each $\boldsymbol{w}_j$ for $j \in [J]$, since it corresponds to the cost function of a rescaled RDCP.

Proposition III.5. For $j \in [J]$, $F_j = f_j + r_j$ is the cost function of RDCP (Def. II.1) for some given $\boldsymbol{x}_j \in \mathbb{R}^N$ and $\boldsymbol{y}_j \in \mathbb{R}^N$. For an initial guess $\boldsymbol{w}_j^{(0)} \in \operatorname{ran}(\mathbf{C}_{\boldsymbol{x}}^\top)$ and the step size $\alpha = L_j/S_{\max}(\boldsymbol{\psi}_j * \boldsymbol{x})$, the iterates $\{\boldsymbol{w}_j^{(t)}\}_{t\in\mathbb{N}}$ of PGD applied to F_j satisfy for every $j \in [J]$:

$$\frac{F_j\big(\boldsymbol{w}^{(t)}\big) - F_j^*}{F_j\big(\boldsymbol{w}^{(0)}\big) - F_j^*} \leq \left(1 - \frac{S^+_{\min}(\boldsymbol{\psi_j} * \boldsymbol{x})}{S_{\max}(\boldsymbol{\psi}_j * \boldsymbol{x})}\right)^t. \tag{17}$$

Proof. For $\boldsymbol{y}_j = L_j^{1/2}(\boldsymbol{\psi}_j * \boldsymbol{y})$ and $\boldsymbol{x}_j = L_j^{-1/2}(\boldsymbol{\psi}_j * \boldsymbol{x})$ we see that $f_j(\boldsymbol{w}_j) = \frac{1}{2L_j}\left\|(\boldsymbol{w}_j * \boldsymbol{\psi}_j * \boldsymbol{x}) - \sqrt{L_j}(\boldsymbol{\psi}_j * \boldsymbol{y})\right\|_2^2 = \frac{1}{2}\left\|\boldsymbol{w}_j * \boldsymbol{x}_j - \boldsymbol{y}_j\right\|_2^2$. Theorem II.14 concludes this result. ☐

Note. Given the result above, we set $L_j = S_{\max}(\boldsymbol{\psi}_j * \boldsymbol{x})$ in order to have a unit step size of $\alpha = 1$.

Algorithm 1 summarizes our method: across the different *subbands* ψ_j, we *normalize adaptively* with respect to the given signal $\boldsymbol{x}$ and derive *kernel estimates* $\boldsymbol{w}_j$ (SNAKE).

Theorem III.6. *Given $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^N$, a Parseval filterbank $\boldsymbol{\Psi} = (\boldsymbol{\psi}_0, \ldots, \boldsymbol{\psi}_{J-1})$ and Multi-RDCP (Def. III.3), the iterates $\{\boldsymbol{W}^{(t+1)}\}_{t\in\mathbb{N}}$ of SNAKE (Algorithm 1) satisfy:*

$$\frac{F_\Psi\big(\mathbf{W}^{(t)}\big) - F_\Psi^*}{F_\Psi\big(\mathbf{W}^{(0)}\big) - F_\Psi^*} \leq \left(1 - \min_{j\in[J]} \frac{S^+_{\min}(\boldsymbol{\psi}_j * \boldsymbol{x})}{S_{\max}(\boldsymbol{\psi}_j * \boldsymbol{x})}\right)^t. \tag{18}$$

Algorithm 1 Subband-Normalized Adaptive Kernel Estimation
Given $\boldsymbol{x}$ and $\boldsymbol{y} = \boldsymbol{h} * \boldsymbol{x} + \boldsymbol{\eta}$ where $\boldsymbol{\eta}$ is noise, finds $\boldsymbol{w} \approx \boldsymbol{h}$.

Require: $(\boldsymbol{\psi}_j)_{j=0}^{J-1}$ (Parseval), $(\lambda_j)_{j=0}^{J-1}$ positive, $T \in \mathbb{N}$

for $j = 0 : J-1$ **do**
  $\boldsymbol{x}_j = \boldsymbol{\psi}_j * \boldsymbol{x}$
  $L_j = S_{\max}(\boldsymbol{x}_j)$ {see Equation (8)}
  $\boldsymbol{\nabla} f_j : \boldsymbol{w}_j \mapsto \frac{1}{L_j}(\boldsymbol{w}_j * \boldsymbol{x}_j - \sqrt{L_j}(\boldsymbol{\psi}_j * \boldsymbol{y})) * \overleftarrow{\boldsymbol{x}}_j$
  $\mathrm{Prox}_{r_j} : \boldsymbol{v} \mapsto \max(0, 1 - \lambda_j / \|\boldsymbol{v}\|_2)\boldsymbol{v}$
  $\boldsymbol{w}_j^{(0)} = \boldsymbol{0}$
  for $t = 0 : T-1$ **do**
    $\boldsymbol{w}_j^{(t+1)} = \mathrm{Prox}_{r_j}\left(\boldsymbol{w}_j^{(t)} - \boldsymbol{\nabla} f_j(\boldsymbol{w}_j^{(t)})\right),$
return $\boldsymbol{w} = \sum_{j=0}^{J-1} \frac{1}{\sqrt{L_j}} \boldsymbol{w}_j^{(T)} * \boldsymbol{\psi}_j * \overleftarrow{\boldsymbol{\psi}}_j$

Proof. We write $F_{\boldsymbol{\Psi}}(\mathbf{W}) = \sum_{j=0}^{J-1} F_j(\boldsymbol{w}_j)$. Proposition III.5 gives us, for each $j \in [J]$:

$$F_j(\boldsymbol{w}_j^{(t)}) - F_j^* \leq \left(1 - \frac{S_{\min}^+(\boldsymbol{\psi}_j * \boldsymbol{x})}{S_{\max}(\boldsymbol{\psi}_j * \boldsymbol{x})}\right)^t \left(F_j(\boldsymbol{w}^{(0)}) - F_j^*\right)$$
$$\leq \left(1 - \min_{j \in [J]} \frac{S_{\min}^+(\boldsymbol{\psi}_j * \boldsymbol{x})}{S_{\max}(\boldsymbol{\psi}_j * \boldsymbol{x})}\right)^t \left(F_j(\boldsymbol{w}^{(0)}) - F_j^*\right).$$

Summing over $j \in [J]$ concludes the result. □

This convergence rate does not depend on the ratio of the smallest to largest positive power spectrum entry $S_{\min}^+(\boldsymbol{x})/S_{\max}^+(\boldsymbol{x})$ like PGD for RDCP, but on the ratios within subbands. This represents an improvement, even with a simple filterbank, as we will demonstrate in the next section.

IV. Numerical simulation

We consider a kernel $\boldsymbol{h} \in \mathbb{R}^N$, $N = 256$ (see Figure 1, left). The input $\boldsymbol{x}$ is a sample of Brownian noise [20]. We observe $\boldsymbol{x}$ and $\boldsymbol{y} = \boldsymbol{h} * \boldsymbol{x} + \boldsymbol{\eta}$, where $\boldsymbol{\eta}$ is an sample of white Gaussian noise with $\|\boldsymbol{\eta}\|_2 = \frac{1}{20}\|\boldsymbol{x}\|_2$. Note that $\boldsymbol{h}$ is narrowband while $\boldsymbol{\eta}$ is broadband. In the absence of regularization, deconvolution of $\boldsymbol{y}$ from $\boldsymbol{x}$ would yield an estimate $\tilde{\boldsymbol{w}}$ such that $\tilde{\boldsymbol{w}} * \boldsymbol{x} \approx \boldsymbol{y}$. Instead, by promoting sparsity across subbands of the filterbank $\boldsymbol{\Psi}$, multi-RDCP aims at $\boldsymbol{w}^*$ such that $\boldsymbol{w}^* \approx \boldsymbol{h}$.

We build a Parseval filterbank $\boldsymbol{\Psi} = (\boldsymbol{\psi}_j)_{j=0}^{J-1}$ of $J = 7$ "real Shannon wavelets" [21]. For every $j \in [J]$ and $k \leq N/2$, $\hat{\boldsymbol{\psi}}_j[k] = 1$ if $2^j < k \leq 2^{j+1}$ and 0 elsewhere. For $k > N/2$, we complete the construction by Hermitian symmetry.

After a preliminary stage of hyperparameter calibration, we select $\lambda = 40\|\boldsymbol{x}\|_2$ for RDCP and $\lambda_j = 150/L_j\|\boldsymbol{x}\|_2$ for Multi-RDCP. We leave the important question of sensitivity to hyperparameters to future work.

The CPU time per iteration is roughly $J = 7$ times greater for SNAKE than for PGD. Yet, this discrepancy is compensated by the faster convergence guarantee of SNAKE: see Theorem III.6. Figure 1 illustrates this phenomenon.

To assess the quality of our reconstructions and justify our choice of regularization, we have employed an oracle benchmark based on Ridge Regression, where the optimal solution $\boldsymbol{w}^*$ is computed with full knowledge of the ground truth $\boldsymbol{h}$. Ridge Regression is defined by the regularizer $r(\boldsymbol{w}) = \lambda\|\boldsymbol{w}\|_2^2$ and since the optimality condition is $(\boldsymbol{w} * \boldsymbol{x} - \boldsymbol{y}) * \overleftarrow{\boldsymbol{x}} + \frac{\lambda}{2}\boldsymbol{w} = 0$, this admits a closed-form solution via the FFT. For each

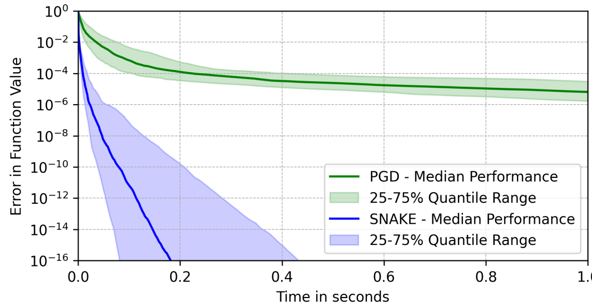

Figure 1. Plot of relative error to optimal cost function value of F (as in LHS of (12)) for PGD (green) and of $F_{\boldsymbol{\Psi}}$ (as in LHS of (18)) for SNAKE (blue) as a function of CPU time. We report the median and quartiles across 100 independent samples of $\boldsymbol{x}$ and $\boldsymbol{\eta}$.

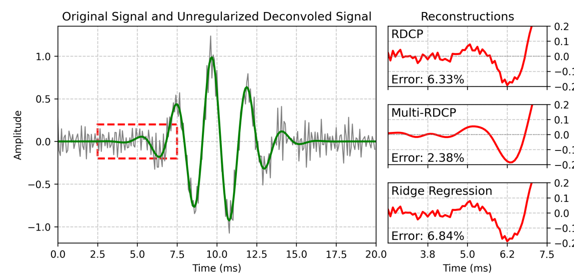

Figure 2. Left: a narrowband signal $\boldsymbol{h}$ (green) and its unregularized deconvolution $\tilde{\boldsymbol{w}}$ (gray) given Brownian noise $\boldsymbol{x}$ and $\boldsymbol{y} = \boldsymbol{h} * \boldsymbol{w} + \boldsymbol{\eta}$ where $\boldsymbol{\eta}$ is white noise. Right: typical solutions $\boldsymbol{w}^*$ or $\boldsymbol{w}^{(T)}$ from regularized deconvolutions. For clarity, we have zoomed onto the red rectangle.

instance of $\boldsymbol{x}$ and $\boldsymbol{\eta}$, we test across different $\lambda > 0$, to see what the best possible $\boldsymbol{w}^*$ achievable by Ridge Regression is, while measuring the reconstruction error as $\|\boldsymbol{w}^* - \boldsymbol{h}\|_2 / \|\boldsymbol{h}\|_2$. Figure 2 illustrates our results.

Over 100 instances, the best possible Ridge Regression reconstruction has a median error of 6.7%. In comparison, the reconstructions of RDCP achieves a median error of 6.7% and Multi-RDCP attains 2.3%.

V. Conclusion

We have presented two contributions: a new regularization technique named multiband robust deconvolution (Multi-RDCP) and a dedicated algorithm named Subband-Normalized Adaptive Kernel Evaluation (SNAKE). SNAKE benefits from filterbank analysis in Multi-RDCP to adapt the step size of subband-wise proximal gradient descents to local spectral characteristics, hence an improved convergence rate. Furthermore, Multi-RDCP promotes sparsity across the subband spectrum of the solution, surpassing the best possible reconstruction performance of Ridge Regression.

Looking ahead, SNAKE's adaptability opens possibilities for other regularization schemes, where classical methods struggle due to high spectral variability. Additionally, exploring adaptive strategies for selecting the subband regularization parameters $(\lambda_j)_{j=0}^{J-1}$ based on prior knowledge of the target signal holds promise to further enhance performance.

REFERENCES

[1] J. Kovacevic, V. K. Goyal, and M. Vetterli, *Fourier and Wavelet Signal Processing*. Cambridge University Press, 2014.

[2] S. M. Riad, "The deconvolution problem: An overview," *Proceedings of the IEEE*, vol. 74, no. 1, pp. 82–85, 1986.

[3] T. Ceccherini-Silberstein, F. Scarabotti, and F. Tolli, *Discrete Harmonic Analysis: Representations, Number Theory, Expanders, and the Fourier Transform*. Cambridge University Press, 2018.

[4] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*, 2nd ed. Upper Saddle River, NJ: Prentice Hall, 1999.

[5] G. Fant, *Speech Acoustics and Phonetics*. Kluwer Academic Publishers, 2004.

[6] N. H. Fletcher and T. D. Rossing, *The physics of musical instruments*. Springer Science & Business Media, 2012.

[7] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 68, no. 1, pp. 49–67, 2006.

[8] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[9] J.-J. Moreau, "Fonctions convexes duales et points proximaux dans un espace hilbertien," *Comptes Rendus de l'Académie des Sciences de Paris*, vol. A255, no. 22, pp. 2897–2899, Nov. 1962.

[10] N. Parikh and S. Boyd, "Proximal algorithms," *Foundations and Trends in Optimization*, vol. 1, no. 3, pp. 127–239, Jan. 2014.

[11] H. Karimi, J. Nutini, and M. Schmidt, "Linear convergence of gradient and proximal-gradient methods under the Polyak–Łojasiewicz condition," in *Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD)*. Springer, 2016, pp. 795–811.

[12] B. T. Polyak, "Gradient methods for the minimisation of functionals," *USSR Computational Mathematics and Mathematical Physics*, vol. 3, no. 4, pp. 864–878, 1963.

[13] S. Łojasiewicz, "Une propriété topologique des sous-ensembles analytiques réels," in *Les équations aux dérivées partielles*, vol. 117. Colloques internationaux du CNRS, 1963, pp. 87–89.

[14] R. A. Horn and C. R. Johnson, *Matrix Analysis*, 2nd ed. Cambridge University Press, 2012.

[15] R. M. Gray, *Toeplitz and Circulant Matrices: A Review*. Foundations and Trends in Communications and Information Theory, 2006, vol. 2, no. 3.

[16] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, ser. Grundlehren der mathematischen Wissenschaften. Berlin: Springer-Verlag, 1998, vol. 317.

[17] P. G. Casazza and G. Kutyniok, *Finite Frames: Theory and Applications*. Springer Science & Business Media, 2012.

[18] D. Haider, F. Perfler, V. Lostanlen, M. Ehler, and P. Balazs, "Hold me tight: Stable encoder-decoder design for speech enhancement," in *Proceedings of the International Speech Communication Association Conference (INTERSPEECH)*. ISCA, 2024.

[19] V. Lostanlen, D. Haider, H. Han, M. Lagrange, P. Balazs, and M. Ehler, "Fitting auditory filterbanks with multiresolution neural networks," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2023.

[20] C. Gardiner, *Stochastic methods*. Springer Berlin Heidelberg, 2009, vol. 4.

[21] M. W. Frazier, *An Introduction to Wavelets through Linear Algebra*. Springer Science & Business Media, 2006.