An approximal operator with application to audio inpainting

Ondřej Mokrý, Pavel Rajmic*

Signal Processing Laboratory, Department of Telecommunications,
Faculty of Electrical Engineering and Communication, Brno University of Technology, Technická 12, 612 00 Brno, Czech Republic

Abstract

Here goes the abstract.

Keywords: proximal operator, proximal algorithms, approximation, sparsity, audio inpainting

1. Introduction

In signal and image processing, the so-called proximal splitting algorithms represent an effective way of finding numerical solutions to various problems [1]. However, despite their conceptual simplicity, proximal algorithms are not always computationally tractable. For instance, in the area of audio signal restoration, it is often necessary to handle the proximal operator of a composition of a linear (or affine) transform and a functional. This happens in many variations of the tasks of audio inpainting [2, 3, 4] (further discussed in Sec. 3), audio declipping [5, 6, 7, 8], and audio dequantization [9, 10, 11].

To introduce the concept more specifically, let $L\colon \mathbb{W} \to \mathbb{V}$ be a linear mapping between two vector spaces, and let $g\colon \mathbb{V} \to \mathbb{R}$ be a convex lower semi-continuous functional. We are interested in computing the proximal operator prox_f where $f = g \circ L$ (the symbol $\circ$ shall denote the composition of two functions), i.e. the mapping L is applied first, followed by g.

In some cases, explicit formulas for $\operatorname{prox}_{g\circ L}$ are available. For instance, [1] provides an explicit formula for the case of real-valued operator L between finite-dimensional spaces such that $LL^\top = \alpha\mathrm{Id}$. A similar result is presented in [12], where L is assumed to be complex-valued and satisfying the condition that LL^* is diagonal (the star denotes the adjoint operator). However, it is limited to the case when prox_g is the operator of projection onto a box-type set.

In the present contribution, we provide analysis of the case when $L^\top L = \alpha\mathrm{Id}$, which is closely related to [1], and it is motivated by a recent signal processing application [4]. In Sec. 2, we derive a formula for $\operatorname{prox}_{g\circ L}$ in such a scenario. Furthermore, an approximation of the derived proximal operator is introduced and analyzed, provided that an explicit formula for $\operatorname{prox}_{\alpha g}$ exists. Sec. 3 shows its usefulness in the case of sparsity-based audio inpainting. The error arising from computing the proximal step of an iterative algorithm only approximately is evaluated on the example of audio inpainting.

*Corresponding author
Email addresses: `170583@vutbr.cz` (Ondřej Mokrý), `rajmic@feec.vutbr.cz` (Pavel Rajmic)

Throughout the whole article, the symbol $\|\cdot\|$ may denote a general norm on a Hilbert space $\mathbb{V}$, the common ℓ_2 norm on finite-dimensional real or complex vector space, or the induced operator norm. The particular case should be clear from the context. The inner product inducing $\|\cdot\|$ will be denoted $\langle\cdot,\cdot\rangle$. Should any other (pseudo)norm appear, it will by identified using the lower index notation, e.g. $\|\cdot\|_0$, $\|\cdot\|_1$.

2. Proximal operator of a composition of a proper convex function with an affine mapping and its approximation

2.1. Theoretical proposition

It was already stated that the novel proposition of the paper is related to the known formula for $\mathrm{prox}_{g\circ L}$ in the case of semi-orthogonal L, i.e. $LL^\top = \alpha\mathrm{Id}$. To build upon this relation, we start with quoting the corresponding lemma.

Lemma 1 (the original lemma from [13, p. 140])**.** *Let $g\colon \mathbb{R}^m \to \mathbb{R}\cup\{\infty\}$ be a proper convex function, and let $f(\mathbf{x}) = g(L\mathbf{x}+\mathbf{b})$, where $\mathbf{b}\in\mathbb{R}^m$ and $L\colon \mathbb{V}\to\mathbb{R}^m$ is a linear transformation satisfying $LL^\top = \alpha\mathrm{Id}$ for some constant $\alpha > 0$. Then for any $\mathbf{x}\in\mathbb{V}$,*

$$\mathrm{prox}_f(\mathbf{x}) = \mathbf{x} + \alpha^{-1}L^\top\left(\mathrm{prox}_{\alpha g}(L\mathbf{x}+\mathbf{b}) - L\mathbf{x} - \mathbf{b}\right). \tag{1}$$

Note that when $\mathrm{prox}_{\alpha g}$ is known explicitly, Eq. (1) provides an explicit form of prox_f. As an example, take $f = \|L\cdot\|_1 = \|\cdot\|_1\circ L$. Then g is the ℓ_1 norm and the corresponding $\mathrm{prox}_{\alpha g}$ is the soft thresholding operator with threshold α. Lemma 1 may be used also when g is the indicator function of a closed convex set $C\subset\mathbb{R}^m$,

$$g(\mathbf{x}) = \iota_C(\mathbf{x}) = \begin{cases} 0 & \mathbf{x}\in C, \\ \infty & \mathbf{x}\notin C. \end{cases} \tag{2}$$

In such a case, the proximal operator of αg is the operator of projection onto C, denoted $\mathrm{prox}_{\alpha g}(\mathbf{x}) = \mathrm{proj}_C(\mathbf{x})$. Additionally, the frame theory provides an example of L that fits the lemma—it may be the synthesis operator of a tight frame [14, 15, 16, 17, 18]. The question is: If we want to employ the *analysis* operator instead, how will formula (1) be affected? The following lemma answers the question.

Lemma 2 (the proposed lemma)**.** *Let $g\colon \mathbb{R}^m \to \mathbb{R}\cup\{\infty\}$ be a proper convex function, and let $f(\mathbf{x}) = g(A\mathbf{x}+\mathbf{b})$, where $\mathbf{b}\in\mathbb{R}^m$ and $A\colon \mathbb{V}\to\mathbb{R}^m$ is a linear transformation satisfying $A^\top A = \alpha\mathrm{Id}$ for some constant $\alpha > 0$. Then for any $\mathbf{x}\in\mathbb{V}$,*

$$\mathrm{prox}_f(\mathbf{x}) = \alpha^{-1}A^\top\left(\mathrm{prox}_{\alpha g+\iota_{(\mathcal{R}(A)+\mathbf{b})}}(A\mathbf{x}+\mathbf{b}) - \mathbf{b}\right), \tag{3}$$

where $\iota_{(\mathcal{R}(A)+\mathbf{b})}$ is the indicator function of the affine space which we obtain by shifting the range space of A by the vector $\mathbf{b}$.

Although the proof is given in Appendix A including the vector $\mathbf{b}$, for simplicity, we will further assume $\mathbf{b}=\mathbf{0}$.

From the viewpoint of frame theory, the operator A in Lemma 2 is the analysis operator of a tight frame. Suppose that $\mathbb{V}=\mathbb{R}^n$, i.e. $A\colon\mathbb{R}^n\to\mathbb{R}^m$. It is a straightforward consequence of the assumption $A^\top A = \alpha\mathrm{Id}$ that in such a case, it must hold $n\le m$ due to the property

$$n = \mathrm{rank}(AA^\top) = \mathrm{rank}(A) \le m. \tag{4}$$

Observe that the crucial difference between the two lemmas is that in the latter case, the image of the proximal operator of αg is forced to lie in the subspace $\mathcal{R}(A)$. Omitting this condition, Eq. (3) would be obtained simply by using Eq. (1) and plugging in the property $A^\top A = \alpha \mathrm{Id}$. Such a step is correct when A is surjective, in which case $m = n$ (as a consequence of Eq. (4)) and the indicator function $\iota_{\mathcal{R}(A)}$ equals zero on whole $\mathbb{R}^n$ and therefore $\mathrm{prox}_{\alpha g + \iota_{\mathcal{R}(A)}} = \mathrm{prox}_{\alpha g}$. However, the consequence of A being full rank[1] is that the operator is unitary and by taking $A = L^\top = L^{-1}$, Lemmas 1 and 2 coincide.

Finally, observe that Lemma 1 provides a constructive way to compute $\mathrm{prox}_{g \circ L}$ when $\mathrm{prox}_{\alpha g}$ is known explicitly. On the contrary, Lemma 2 is not constructive in the case of $n < m$, since the explicit form of $\mathrm{prox}_{\alpha g}$ does not suffice to have an explicit form of $\mathrm{prox}_{\alpha g + \iota_{\mathcal{R}(A)}}$.

2.2. Explicit approximation

The strength of Lemma 1 is that it offers an explicit form of prox_f when the explicit form of $\mathrm{prox}_{\alpha g}$ is available. This is possible for example in the aforementioned cases of (weighted) ℓ_1 norm or the indicator function in the role of g. However, the proximal operator including an additional restriction, which is the case of $\mathrm{prox}_{\alpha g + \iota_{\mathcal{R}(A)}}$ in Lemma 2, is seldom known[2], resulting in the need for a reasonable approximation of $\mathrm{prox}_{\alpha g + \iota_{\mathcal{R}(A)}}$.

Since the orthogonal projection onto $\mathcal{R}(A)$ in the case of $A^\top A = \alpha \mathrm{Id}$ is expressed easily as [14]

$$\mathrm{proj}_{\mathcal{R}(A)}(\mathbf{x}) = \alpha^{-1} A A^\top \mathbf{x}, \tag{5}$$

a natural possibility is to study the approximation

$$\mathrm{prox}_{\alpha g + \iota_{\mathcal{R}(A)}} \approx \mathrm{proj}_{\mathcal{R}(A)} \circ \mathrm{prox}_{\alpha g}, \tag{6}$$

i.e. the composition of two known operators. Such an approximation will thus take the form of

$$\mathrm{approx}_f(\mathbf{x}) = \alpha^{-1} A^\top \left(\mathrm{proj}_{\mathcal{R}(A)} \left(\mathrm{prox}_{\alpha g}(A\mathbf{x}) \right) \right). \tag{7}$$

where we introduced the denotation approx for the *approximal operator*.

The most important property of approx_f is pointed out in the following lemma.

Lemma 3. *Under the conditions of Lemma 2, the approximal operator is*

$$\mathrm{approx}_f(\mathbf{x}) = \alpha^{-1} A^\top \mathrm{prox}_{\alpha g}(A\mathbf{x}). \tag{8}$$

and it is a proximal operator of some convex lower semi-continuous function.

Note that the shortened form of approx_f in Eq. (8) was obtained by plugging (5) into (7) and recalling the property $A^\top A = \alpha \mathrm{Id}$.

The lemma is proven in Appendix A, justifying two properties of proximal operators, one of which is the non-expansivity of approx_f. As mentioned in [19, p. 7], the non-expansivity plays a role in convergence analysis of iterative proximal algorithms. However, it should be emphasized that the lemma is not constructive and that it is unclear which (convex lower semi-continuous) function in particular this proximal operator corresponds to.

In the following section, an example from the field of audio processing demonstrates the suitability of approx_f as an approximation of prox_f.

---

[1] In the case of $A\colon \mathbb{R}^n \to \mathbb{R}^m$, the operator A has full rank if $\mathrm{rank}(A) = \max\{m, n\}$.

[2] The rare example is the proximal operator of (weighted) ℓ_1 norm over a box [13, pp. 145–146].

3. Experiments

3.1. Sparsity-based audio inpainting

Audio inpainting is a rather modern term for the task of filling highly degraded or missing samples of digital audio [3]. The same task is referred to as the interpolation of missing samples [2, 20] or packet-loss concealment [21, 22].

Popular audio inpainting methods are based on the assumption that musical audio signals are sparse with respect to a suitable time-frequency (TF) transform. To proceed with formalization of this assumption, denote $\mathbf{y} \in \mathbb{C}^n$ the observed (i.e., degraded) signal and let $M\colon \mathbb{C}^n \to \mathbb{C}^n$ be the operator which fills the missing information in the signal with zeros. Denote Γ the set of signals consistent with the observed signal $\mathbf{y}$,

$$\Gamma = \{\mathbf{x} \in \mathbb{C}^n \mid M\mathbf{x} = M\mathbf{y}\}\,. \tag{9}$$

The inpainting task is then formulated as the following optimization task:

Find the signal from Γ with the corresponding TF coefficients as sparse as possible. (10)

Let us recall two operators: Let $A\colon \mathbb{C}^n \to \mathbb{C}^m$, $m \geq n$, be the analysis operator, expanding the time-domain signal into the vector of TF coefficients, and let $A^*\colon \mathbb{C}^m \to \mathbb{C}^n$ be its synthesis counterpart, producing a signal, given the coefficients. With this notation, (10) can be understood in two ways:

$$\operatorname*{arg\,min}_{\mathbf{z}} \ \|\mathbf{z}\|_0 \quad \text{s.t.} \quad A^*\mathbf{z} \in \Gamma, \tag{11a}$$

$$\operatorname*{arg\,min}_{\mathbf{x}} \ \|A\mathbf{x}\|_0 \quad \text{s.t.} \quad \mathbf{x} \in \Gamma. \tag{11b}$$

The symbol $\|\cdot\|_0$ denotes sparsity, i.e. the pseudonorm that counts the non-zero entries of the argument. Since Eq. (11a) includes the synthesis operator, it is referred to as the *synthesis formulation*; by the same reasoning, Eq. (11b) is called the *analysis formulation* [23].

Note that the two formulations are equivalent only when the operator A is unitary, i.e. it holds $A^* = A^{-1}$. However, this is not the case of the commonly used Gabor, wavelet or ERBlet transforms [15, 18].

3.1.1. Convex relaxation

Formulations (11) are problematic in that they include the ℓ_0 pseudonorm, resulting in the task being NP-hard. Two possible classes of methods exist to solve (11), in general only approximately. Either a non-convex heuristic algorithm is employed to tackle (11), e.g. the OMP [3] or SPAIN [24]. Alternatively, the task needs to be relaxed to a convex optimization problem [25, 26]. Denoting $\mathcal{S}\colon \mathbb{C}^m \to \mathbb{R}$ a convex sparsity-related penalty, the convex relaxations of (11a) and (11b) attain the form

$$\operatorname*{arg\,min}_{\mathbf{z}} \ \{\mathcal{S}(\mathbf{z}) + \iota_\Gamma(A^*\mathbf{z})\}\,, \tag{12a}$$

$$\operatorname*{arg\,min}_{\mathbf{x}} \ \{\mathcal{S}(A\mathbf{x}) + \iota_\Gamma(\mathbf{x})\}\,. \tag{12b}$$

The constraints from (11) are now included in the objective function using the indicator function ι_Γ. Note that Γ is a convex set by design. Thus, in order to have an overall convex problem, $\mathcal{S}$ has to be a convex function, and it shall promote sparsity. As an example of a suitable and widely used penalty $\mathcal{S}$, we mention the (weighted) ℓ_1 norm.

3.1.2. Inconsistent reformulation

In (12), the solution is forced to be equal to the observed signal in its reliable (non-distorted) parts. However, this assumption may be too strong, for instance when the observed signal $\mathbf{y}$ is noisy. In such a case, the alternative reformulations to solve are the so-called inconsistent problems

$$\arg\min_{\mathbf{z}} \; \{\mathcal{S}(\mathbf{z}) + \lambda\|MA^*\mathbf{z} - M\mathbf{y}\|\}, \tag{13a}$$

$$\arg\min_{\mathbf{x}} \; \{\mathcal{S}(A\mathbf{x}) + \lambda\|M\mathbf{x} - M\mathbf{y}\|\}, \tag{13b}$$

where the parameter $\lambda > 0$ balances consistency with the data and the sparsity.

3.1.3. Solving the task

Formulations (12) and (13) consist of sums of lower semicontinuous convex functions. Such a scenario allows the use of proximal algorithms to solve the tasks numerically [1]. To find the minimum of a sum of lower semicontinuous convex functions f_1 and f_2 using the proximal splitting approach, one must be able to evaluate both prox_{f_1} and prox_{f_2}, or, in the case of differentiable f_1 or f_2, the corresponding gradient. These are evaluated in every iteration, meaning that simple, computationally cheap formulas are preferred.

To choose suitable algorithms for solving (12) and (13), suppose that the explicit form of $\mathrm{prox}_{\mathcal{S}}$ is available. Note that in practice, this assumption is not too restricting. Sometimes the situation is even the opposite, meaning that a suitable operator in the place of $\mathrm{prox}_{\mathcal{S}}$ is used although an explicit form of $\mathcal{S}$ is not available, for instance in the case of (persistent) empirical Wiener and similar operators [27, 19]. Moreover, assume using a tight frame as the TF transformation, i.e. $AA^* = \alpha\mathrm{Id}$. To solve the synthesis model (12a), Lemma 1 is used to compute the prox of the second term,

$$\mathrm{prox}_{\iota_\Gamma \circ A^*}(\mathbf{x}) = \mathbf{x} + \alpha^{-1}A\left(\mathrm{proj}_\Gamma(A^*\mathbf{x}) - A^*\mathbf{x}\right), \tag{14}$$

which enables the use of the Douglas–Rachford (DR) algorithm [1, Sec. IV]. The reason is that for any time-domain signal $\mathbf{s} \in \mathbb{C}^n$, the projection $\mathrm{proj}_\Gamma(\mathbf{s})$ is computed simply entry-by-entry by setting the samples at the reliable positions equal to the observed ones, and keeping the rest.

It should be pointed out that Lemma 1 was used here to handle a complex-valued operator. This is treated in Appendix B, where one can find the argumentation why using complex variables in place of the real variables does not affect the proposed theoretical results of both lemmas.

To solve the analysis model (12b), $\mathrm{prox}_{\mathcal{S}\circ A}$ has to be known for the use in the DR algorithm; formally, it is constructed using Lemma 2 as

$$\mathrm{prox}_{\mathcal{S}\circ A}(\mathbf{x}) = \alpha^{-1}A^*\mathrm{prox}_{\alpha\mathcal{S}+\iota_{\mathcal{R}(A)}}(A\mathbf{x}). \tag{15}$$

Numerical treatment of (15), however, requires choosing one of the following three options:

- $\mathrm{prox}_{\alpha\mathcal{S}+\iota_{\mathcal{R}(A)}}$ is only aproximated by the operator $\mathrm{approx}_{\mathcal{S}\circ A}$ defined by Eq. (8),
- $\mathrm{prox}_{\mathcal{S}\circ A}(\mathbf{x}) = \arg\min_{\mathbf{u}} \left\{\mathcal{S}(A\mathbf{u}) + \frac{1}{2}\|\mathbf{u} - \mathbf{x}\|^2\right\}$ is computed using a nested iterative subroutine, since it is from definition nothing but another optimization task,
- finally, the Chambolle–Pock (CP) algorithm [28] can also be used instead of the DR algorithm to solve (12b) without nested iterations, since it allows composing a convex functional with a linear mapping.

In analogy to the above, solving (13a) is not a problem, but for (13b), the approximation of $\mathrm{prox}_{\mathcal{S}\circ A}$ is needed. For both inconsistent cases, a suitable iterative algorithm to produce the numerical result is ISTA or FISTA (fast iterative shrinkage-thresholding algorithm) [29].

3.2. Evaluation setting

Our experiments basically follow and extend the experiment from [4, Sec. 3]. The motivation is that [4] applies Lemma 1 instead of Lemma 2 to compute $\mathrm{prox}_{\mathcal{S}\circ A}$. In turn, [4] actually uses $\mathrm{approx}_{\mathcal{S}\circ A}$ (unintentionally). It is clear from plugging A in the place of L in Eq. (1), which (recalling $A^\top A = \alpha\mathrm{Id}$) directly produces $\mathrm{approx}_{\mathcal{S}\circ A}$. The MATLAB codes and data from [4] are available at `https://github.com/flieb/AudioInpainting`. We recomputed their results using the shared code, and above that, we aimed at answer the question: as a result ?

answering

Does the more accurate (yet more complicated) use of $\mathrm{prox}_{\mathcal{S}\circ A}$ *produce better results in the audio inpainting task, compared to the use of* $\mathrm{approx}_{\mathcal{S}\circ A}$*?*

Thus, our contribution regards the analysis model. There we decided to use the CP algorithm instead of the DR algorithm in the consistent case, and in the inconsistent case, the CP algorithm approximates $\mathrm{prox}_{\mathcal{S}\circ A}$ in each iteration of FISTA.

Exactly as in [4], the algorithms are tested on four musical signals from [30] sampled at 44.1 kHz that are distorted by dropping out 80 % of the samples at random positions. As the convex sparsity-related penalty, the ℓ_1 norm is used, i.e. $\mathcal{S}(\cdot) = \|\cdot\|_1$. Three different transforms [4, Sec. 3.3.1] are used:

- Gabor transform (GAB) with Hann window of length 23 ms (1024 samples), time sampling parameter $a = 3.6$ ms and $M = 3125$ frequency channels,
- ERBlet transform (ERB) with $qvar = 0.08$ and $bins = 18$,
- wavelet transform (WAV) with $f_{\min} = 100$ Hz, $bw = 3$ Hz and $bins = 120$.

The reconstruction performance is evaluated using the SNR, computed in consistence with [4] as

$$\mathrm{SNR}(\mathbf{s}, \hat{\mathbf{s}}) = 20 \log_{10} \frac{\sigma(\mathbf{s})}{\sigma(\mathbf{s} - \hat{\mathbf{s}})}, \tag{16}$$

where $\mathbf{s}$ is the original (undistorted) signal, $\hat{\mathbf{s}}$ is the reconstructed signal and σ denotes the standard deviation. It is worth noting that the reliable parts of $\mathbf{s}$ and $\hat{\mathbf{s}}$ are not taken into account in Eq. (16).

3.3. Results and discussion

Tab. 1 is a slightly reordered reproduction of the table provided in [4, Sec. 4]. As explained above, the results based on the synthesis model correspond to (12a) and (13a), whereas the analysis-based results only approximate the solutions to Eq. (12b) and (13b), since the operator $\mathrm{approx}_{\mathcal{S}\circ A}$ is used.

Since our modifications stem from the use of $\mathrm{approx}_{\mathcal{S}\circ A}$ in the analysis-based model, Tab. 2 presents only the results of this model. Also, for better readability, only the difference between the values of SNR is shown.

Tab. 2 shows two remarkable results. On the right side, the SNR for FISTA is almost independent of whether $\mathrm{prox}_{\mathcal{S}\circ A}$ is computed accurately or not. Second note is to the left side, showing results of DR and CP algorithm, representing the use of $\mathrm{approx}_{\mathcal{S}\circ A}$ and $\mathrm{prox}_{\mathcal{S}\circ A}$, respectively. Here, the more accurate approach with $\mathrm{prox}_{\mathcal{S}\circ A}$ outperforms the approximate approach with $\mathrm{approx}_{\mathcal{S}\circ A}$ only

Table 1: Values of SNR in dB from [4, Tab. 2], based on the four test signals, three TF dictionaries and both for the consistent and inconsistent approach, corresponding to problems (12) and (13), respectively.

#		DR (consistent)			FISTA (inconsistent)		
		GAB	WAV	ERB	GAB	WAV	ERB
1	Synthesis	18.7	26.0	26.4	15.5	25.5	25.9
	Analysis	16.9	25.9	26.3	18.6	25.2	25.6
2	Synthesis	20.1	25.9	25.9	16.8	25.1	25.2
	Analysis	18.3	25.7	25.6	19.7	25.1	25.2
3	Synthesis	18.6	19.2	19.3	17.4	18.9	19.1
	Analysis	17.9	19.2	19.3	18.5	19.2	19.2
4	Synthesis	16.2	19.8	20.4	13.6	19.3	20.1
	Analysis	15.1	19.7	20.4	16.1	19.7	20.4

Table 2: Difference between the values of SNR taken from the original experiment and from its correct implementation. The latter one uses the CP algorithm instead of the DR algorithm in the consistent case, and in the inconsistent case, $\mathrm{prox}_{\mathcal{S}\circ A}$ is evaluated using the CP algorithm in every iteration of FISTA. Only the results of the analysis-based approach are shown. Positive values indicate cases in which the new implementation performs better.

#	DR/CP (consistent)			FISTA (inconsistent)		
	GAB	WAV	ERB	GAB	WAV	ERB
1	1.71	−0.28	−0.61	0.02	0.01	0.01
2	1.72	−0.13	−0.24	0.11	−0.02	0.00
3	0.68	−0.05	−0.02	0.03	0.00	−0.02
4	1.11	0.07	0.00	0.02	−0.01	−0.01

with the Gabor dictionary, whereas it performs slightly worse for wavelets and ERBlets in most cases.

Recall that the inconsistent analysis model employs a nested iterative CP algorithm within FISTA, which results in the computational cost being remarkably higher compared to the synthesis model.

Furthermore, we examine the performance of the DR (original) and CP (new) algorithms in the analysis model while choosing more strict convergence criteria than in the previously described experiment. In the implementation of [4], the parameters indicating convergence are the maximum number of iterations (`param.maxit`) and the relative norm of solutions in subsequent iterations (`param.tol`). The algorithm stops when either of the criteria is reached. The settings of these parameters in the experiments are as follows:

criterion	setting for Tables 1 and 2	setting for Tables 3 and 4
`param.maxit`	200	500
`param.tol`	10^{-3}	10^{-5}

In Tab. 3, also the results of synthesis-based model using the DR algorithm are recomputed with the new choice of parameters. Note that the inconsistent approach is omitted, since even with the less strict convergence criteria, the results of the two approaches were almost indistinguishable. For the purpose of direct comparison, Tab. 4 presents the differences of SNR shown in Tab. 3. As above, positive values indicate better performance of the implementation based on $\mathrm{prox}_{\mathcal{S}\circ A}$.

Table 3: Values for the DR (original) and the CP (new) algorithms for the analysis-based approach. Parameters were set to `param.maxit = 500` and `param.tol = 1e-5`.

#	DR			CP		
	GAB	WAV	ERB	GAB	WAV	ERB
1	18.01	26.44	26.79	18.52	26.52	26.84
2	19.34	26.23	26.28	19.98	26.19	26.25
3	18.39	19.28	19.37	18.58	19.24	19.36
4	15.84	19.77	20.45	16.14	19.75	20.42

Table 4: Difference of the values of SNR in dB presented in Tab. 3.

DR/CP		
GAB	WAV	ERB
0.51	0.08	0.05
0.64	−0.04	−0.03
0.19	−0.04	−0.01
0.30	−0.02	−0.03

It is clear from Tab. 4 that letting the algorithms converge closer to the actual solution of corresponding optimization tasks reduces the difference between the more and less accurate approaches. Nonetheless, it can still be seen that not only the results in general but also the inaccuracy of DR algorithm in the analysis-based approach depend on the choice of (redundant) time-frequency representation of the audio signal.

4. Conclusion

- here goes the conclusion

Although only a short experiment on audio inpainting was presented, the theoretical result has a straightforward application also to the problems of audio declipping or dequantization, or even in image processing.

Appendix A. Proofs

Proof of Lemma 2. We start the proof by quoting the first part of the proof of Lemma 1, as presented in [13, pp. 140–141], since the assumed property of the linear transform is not crucial at first.

By definition, $\mathrm{prox}_f\colon \mathbb{V} \to \mathbb{V}$ for $f(\mathbf{u}) = g(A\mathbf{u} + \mathbf{b})$ is a mapping such that $\mathrm{prox}_f(\mathbf{x})$ is the optimal solution of

$$\min_{\mathbf{u}\in\mathbb{V}} \left\{ g(A\mathbf{u} + \mathbf{b}) + \frac{1}{2}\|\mathbf{u} - \mathbf{x}\|^2 \right\}. \tag{A.1}$$

The above problem can be formulated as the following constrained problem:

$$\begin{array}{ll} \min_{\mathbf{u}\in\mathbb{V},\mathbf{z}\in\mathbb{R}^m} & \left\{ g(\mathbf{z}) + \frac{1}{2}\|\mathbf{u} - \mathbf{x}\|^2 \right\} \\ \text{s. t.} & \mathbf{z} = A\mathbf{u} + \mathbf{b}. \end{array} \tag{A.2}$$

Denote the optimal solution of (A.2) by $(\tilde{\mathbf{u}}, \tilde{\mathbf{z}})$ (the existence and uniqueness of $\tilde{\mathbf{u}}$ and $\tilde{\mathbf{z}}$ follow from the underlying assumption that g is proper close and convex). Note that $\tilde{\mathbf{u}} = \mathrm{prox}_f(\mathbf{x})$. Fixing $\mathbf{z} = \tilde{\mathbf{z}}$, we obtain that $\tilde{\mathbf{u}}$ is the optimal solution of

$$\begin{array}{ll} \min_{\mathbf{u}\in\mathbb{V}} & \frac{1}{2}\|\mathbf{u} - \mathbf{x}\|^2 \\ \text{s. t.} & A\mathbf{u} = \tilde{\mathbf{z}} - \mathbf{b}. \end{array} \tag{A.3}$$

Since strong duality holds for the problem (A.3) [13, pp. 439–440], it follows that there exists $\mathbf{y} \in \mathbb{R}^m$ for which the two conditions

$$\tilde{\mathbf{u}} \in \arg\min_{\mathbf{u}\in\mathbb{V}} \left\{ \frac{1}{2}\|\mathbf{u}-\mathbf{x}\|^2 + \langle \mathbf{y}, A\mathbf{u} - \tilde{\mathbf{z}} + \mathbf{b} \rangle \right\}, \tag{A.4}$$

$$A\tilde{\mathbf{u}} = \tilde{\mathbf{z}} - \mathbf{b}. \tag{A.5}$$

are satisfied. Since the objective function in (A.4) is strictly convex and differentiable, the unique minimizer $\tilde{\mathbf{u}}$ is obtained by setting its gradient to zero, which leads to

$$\tilde{\mathbf{u}} = \mathbf{x} - A^\top \mathbf{y}. \tag{A.6}$$

Substituting this expression of $\tilde{\mathbf{u}}$ into (A.5), we obtain

$$A(\mathbf{x} - A^\top \mathbf{y}) = \tilde{\mathbf{z}} - \mathbf{b}. \tag{A.7}$$

Quoting [13] must be finished here, since we will further utilize different assumption on the linear operator.

Since $A^\top A = \alpha\mathrm{Id}$, applying synthesis $A^\top$ onto both sides of Eq. (A.7) leads to

$$\mathbf{x} - A^\top \mathbf{y} = \alpha^{-1} A^\top (\tilde{\mathbf{z}} - \mathbf{b}). \tag{A.8}$$

The important observation here is that (A.8) is equivalent to (A.7) only when $\tilde{\mathbf{z}} - \mathbf{b} \in \mathcal{R}(A)$, which is enforced by the left hand side of the relation (A.7). Substituting this result into (A.6) leads to an explicit expression for $\tilde{\mathbf{u}}$ in terms of $\tilde{\mathbf{z}}$ (still limited to $\tilde{\mathbf{z}} - \mathbf{b} \in \mathcal{R}(A)$):

$$\tilde{\mathbf{u}} = \alpha^{-1} A^\top (\tilde{\mathbf{z}} - \mathbf{b}). \tag{A.9}$$

Plugging this result in the minimization problem (A.2), we obtain that $\tilde{\mathbf{z}}$ is given by

$$\tilde{\mathbf{z}} = \arg\min_{\mathbf{z}} \left\{ g(\mathbf{z}) + \iota_{(\mathcal{R}(A)+\mathbf{b})}(\mathbf{z}) + \frac{1}{2} \left\| \alpha^{-1} A^\top (\mathbf{z} - \mathbf{b}) - \mathbf{x} \right\|^2 \right\}, \tag{A.10}$$

where the relation $\tilde{\mathbf{z}} - \mathbf{b} \in \mathcal{R}(A)$ is enforced by the indicator function $\iota_{(\mathcal{R}(A)+\mathbf{b})}$.

Now recall two useful properties:

1. Viewing A from the perspective of frame theory, it is an analysis operator corresponding to a tight frame, for which it holds $\|A\mathbf{z}\|^2 = \alpha\|\mathbf{z}\|^2$ for all $\mathbf{z} \in \mathbb{V}$.
2. Since we require $\tilde{\mathbf{z}} - \mathbf{b} \in \mathcal{R}(A)$, it follows from Eq. (5) that $\alpha^{-1} A A^\top (\tilde{\mathbf{z}} - \mathbf{b}) = \tilde{\mathbf{z}} - \mathbf{b}$.

Using the first property, we can rewrite (A.10) as

$$\tilde{\mathbf{z}} = \arg\min_{\mathbf{z}} \left\{ g(\mathbf{z}) + \iota_{(\mathcal{R}(A)+\mathbf{b})}(\mathbf{z}) + \frac{1}{2\alpha} \left\| A\alpha^{-1} A^\top (\mathbf{z} - \mathbf{b}) - A\mathbf{x} \right\|^2 \right\}. \tag{A.11}$$

Using the second property and multiplying the objective function by the (positive) constant α leads to

$$\tilde{\mathbf{z}} = \arg\min_{\mathbf{z}} \left\{ \alpha g(\mathbf{z}) + \iota_{(\mathcal{R}(A)+\mathbf{b})}(\mathbf{z}) + \frac{1}{2} \left\| \mathbf{z} - (A\mathbf{x} + \mathbf{b}) \right\|^2 \right\} \tag{A.12}$$

$$= \mathrm{prox}_{\alpha g + \iota_{(\mathcal{R}(A)+\mathbf{b})}} (A\mathbf{x} + \mathbf{b}). \tag{A.13}$$

Plugging the expression for $\tilde{\mathbf{z}}$ into (A.9) produces the desired result. □

Proof of Lemma 3. To prove that a function $F\colon \mathbb{V} \to \mathbb{V}$ is a proximal operator of a convex lower semi-continuous function, it is sufficient to show two properties [31, Corollary 10.c]:

1. there exists a convex lower semi-continuous function ψ such that for any $\mathbf{y} \in \mathbb{V}$, $F(\mathbf{y}) \in \partial\psi(\mathbf{y})$,
2. F is non-expansive, i.e.

$$\|F(\mathbf{y}) - F(\mathbf{y}')\| \leq \|\mathbf{y} - \mathbf{y}'\|, \quad \forall \mathbf{y}, \mathbf{y}' \in \mathbb{V}. \tag{A.14}$$

The symbol $\partial\psi(\mathbf{y})$ denotes the subdifferential of function ψ at point $\mathbf{y}$, see e.g. [32, Def. 1.2.1, p. 167].

Observe that the first property is necessarily satisfied for $F = \mathrm{prox}_{\alpha g}$, since it is a proximal operator. This means that there exists a convex lower semi-continuous function η such that

$$\forall \mathbf{x} \in \mathbb{R}^m \quad \mathrm{prox}_{\alpha g}(\mathbf{x}) \in \partial\eta(\mathbf{x}) = \{\mathbf{s} \mid \eta(\mathbf{y}) \geq \eta(\mathbf{x}) + \langle \mathbf{s}, \mathbf{y} - \mathbf{x}\rangle \ \forall \mathbf{y} \in \mathbb{R}^m\}. \tag{A.15}$$

If this relation holds for all $\mathbf{x}$, it must hold also for vectors in the form $\mathbf{x} = A\mathbf{z}$, leading to

$$\forall \mathbf{z} \in \mathbb{V} \quad \mathrm{prox}_{\alpha g}(A\mathbf{z}) \in \partial\eta(A\mathbf{z}) = \{\mathbf{s} \mid \eta(\mathbf{y}) \geq \eta(A\mathbf{z}) + \langle \mathbf{s}, \mathbf{y} - A\mathbf{z}\rangle \ \forall \mathbf{y} \in \mathbb{R}^m\}. \tag{A.16}$$

Then also

$$\forall \mathbf{z} \quad \alpha^{-1}A^\top \mathrm{prox}_{\alpha g}(A\mathbf{z}) \in \alpha^{-1}A^\top \partial\eta(A\mathbf{z}) \tag{A.17}$$

which, recalling the property [32, Theorem 4.2.1, p. 184]

$$\partial(h \circ A)(\mathbf{x}) = A^\top \partial h(A\mathbf{x}) \quad \forall \mathbf{x} \in \mathbb{R}^n, \tag{A.18}$$

implies

$$\forall \mathbf{z} \quad \alpha^{-1}A^\top \mathrm{prox}_{\alpha g}(A\mathbf{z}) \in \alpha^{-1}\partial(\eta \circ A)(\mathbf{z}) = \partial(\alpha^{-1}\eta \circ A)(\mathbf{z}). \tag{A.19}$$

Since η is a lower semi-continuous and convex function and A is a linear operator, also $\eta \circ A$ and $\alpha^{-1}\eta \circ A$ are lower semi-continuous and convex. Property 1. is thus satisfied for $F = \mathrm{approx}_f$ with $\psi = \alpha^{-1}\eta \circ A$.

The non-expansiveness can be shown similarly: Substituting (8) into (A.14) and using fundamental property of operator norm leads to

$$\|\alpha^{-1}A^\top \left(\mathrm{prox}_{\alpha g}(A\mathbf{y}) - \mathrm{prox}_{\alpha g}(A\mathbf{y}')\right)\| \leq \|\alpha^{-1}A^\top\| \|\mathrm{prox}_{\alpha g}(A\mathbf{y}) - \mathrm{prox}_{\alpha g}(A\mathbf{y}')\|. \tag{A.20}$$

First, let us compute the operator norm of $\|\alpha^{-1}A^\top\|$. The property $\|A\mathbf{x}\|^2 = \alpha\|\mathbf{x}\|^2 \ \forall \mathbf{x}$ implies $\|A\mathbf{x}\| = \sqrt{\alpha}\|\mathbf{x}\|$, i.e. $\|A\| = \sqrt{\alpha}$. Thus

$$\|\alpha^{-1}A^\top\| = \alpha^{-1}\|A^\top\| = \alpha^{-1}\|A\| = \frac{1}{\sqrt{\alpha}}. \tag{A.21}$$

Now, since $\mathrm{prox}_{\alpha g}$ meets (A.14), it holds

$$\|\mathrm{prox}_{\alpha g}(A\mathbf{y}) - \mathrm{prox}_{\alpha g}(A\mathbf{y}')\| \leq \|A\mathbf{y} - A\mathbf{y}'\| = \|A(\mathbf{y} - \mathbf{y}')\| = \sqrt{\alpha}\|\mathbf{y} - \mathbf{y}'\|. \tag{A.22}$$

Plugging (A.21) and (A.22) into (A.20) shows that (A.14) truly holds for the approx_f operator. $\square$

Appendix B. Comments on the use of complex-valued operators

In the experimental part, complex-valued analysis operator $A\colon \mathbb{C}^n \to \mathbb{C}^m$ and its synthesis counterpart A^* (instead of $A^\top$) are used. Nonetheless, both Lemmas 1 and 2 remain valid also in such a case. The place in the proof that could make trouble in the complex case is the formulation of the Lagrangian in Eq. (A.4), since it should be real; the rest of the manipulations hold also in the complex case.

To deal with the Lagrangian for complex variables, observe that

$$A\mathbf{u} = \tilde{\mathbf{z}} - \mathbf{b} \iff \Re(A\mathbf{u} - \tilde{\mathbf{z}} + \mathbf{b}) = \mathbf{0} \wedge \Im(A\mathbf{u} - \tilde{\mathbf{z}} + \mathbf{b}) = \mathbf{0}, \tag{B.1}$$

which can be rewritten in matrix form as

$$\underbrace{\begin{bmatrix} \Re(A) & -\Im(A) \\ \Im(A) & \Re(A) \end{bmatrix}}_{\hat{A}} \underbrace{\begin{bmatrix} \Re(\mathbf{u}) \\ \Im(\mathbf{u}) \end{bmatrix}}_{\hat{\mathbf{u}}} - \underbrace{\begin{bmatrix} \Re(\tilde{\mathbf{z}} - \mathbf{b}) \\ \Im(\tilde{\mathbf{z}} - \mathbf{b}) \end{bmatrix}}_{\hat{\mathbf{c}}} = \mathbf{0}, \tag{B.2}$$

where $\hat{A}\colon \mathbb{R}^{2n} \to \mathbb{R}^{2m}$, $\hat{\mathbf{u}} \in \mathbb{R}^{2n}$ and $\hat{\mathbf{c}} \in \mathbb{R}^{2m}$. Rewriting $\mathbf{x} \in \mathbb{C}^n$ in a similar way into a real vector $\hat{\mathbf{x}} \in \mathbb{R}^{2n}$ leads to the Lagrangian

$$\frac{1}{2}\|\hat{\mathbf{u}} - \hat{\mathbf{x}}\|^2 + \langle \hat{\mathbf{y}}, \hat{A}\hat{\mathbf{u}} - \hat{\mathbf{c}} \rangle, \tag{B.3}$$

where $\hat{\mathbf{y}} \in \mathbb{R}^{2m}$. The unique minimizer of (B.3) is

$$\hat{\mathbf{u}} = \hat{\mathbf{x}} - \hat{A}^\top \hat{\mathbf{y}}. \tag{B.4}$$

It can be easily shown that the real and imaginary parts of the complex solution $\tilde{\mathbf{u}} = \mathbf{x} - A^*\mathbf{y}$ obtained using Eq. (A.6) match the results in just presented real case.

Note that in the case of the complex Gabor transform presented beforehand, the mapping prox_f should map the real signal $\mathbf{x}$ to a real signal $\tilde{\mathbf{u}}$. This would be ensured by a particular convex-conjugate structure of A, A^* and also $\mathbf{y}$ in such a case. For more detailed description, see the discussion in [12, p. 9] and references therein.

Similar argumentation would be used regarding proof of Lemma 3 and the concept of subdifferentials, where once again, real inner product of complex vectors would be defined.

References

[1] P. Combettes, J. Pesquet, Proximal splitting methods in signal processing, Fixed-Point Algorithms for Inverse Problems in Science and Engineering (2011) 185–212doi:10.1007/978-1-4419-9569-8_10.

[2] A. J. E. M. Janssen, R. N. J. Veldhuis, L. B. Vries, Adaptive interpolation of discrete-time signals that can be modeled as autoregressive processes, IEEE Trans. Acoustics, Speech and Signal Processing 34 (2) (1986) 317–330. doi:10.1109/TASSP.1986.1164824.

[3] A. Adler, V. Emiya, M. Jafari, M. Elad, R. Gribonval, M. Plumbley, Audio Inpainting, IEEE Transactions on Audio, Speech, and Language Processing 20 (3) (2012) 922–932. doi:10.1109/TASL.2011.2168211.

[4] F. Lieb, H.-G. Stark, Audio inpainting: Evaluation of time-frequency representations and structured sparsity approaches, Signal Processing 153 (2018) 291–299. doi:10.1016/j.sigpro.2018.07.012.

[5] K. Siedenburg, M. Kowalski, M. Dorfler, Audio declipping with social sparsity, in: Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on, IEEE, 2014, pp. 1577–1581.

[6] P. Záviška, P. Rajmic, J. Schimmel, Psychoacoustically motivated audio declipping based on weighted l1 minimization, in: 2019 42nd International Conference on Telecommunications and Signal Processing (TSP), 2019, pp. 338–342. doi:10.1109/TSP.2019.8769109.

[7] S. Kitić, L. Jacques, N. Madhu, M. Hopwood, A. Spriet, C. De Vleeschouwer, Consistent iterative hard thresholding for signal declipping, in: Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on, 2013, pp. 5939–5943. doi:10.1109/ICASSP.2013.6638804.

[8] S. Kitić, N. Bertin, R. Gribonval, Sparsity and cosparsity for audio declipping: a flexible nonconvex approach, in: LVA/ICA 2015 – The 12th International Conference on Latent Variable Analysis and Signal Separation, Liberec, Czech Republic, 2015.

[9] L. Rencker, F. Bach, W. Wang, M. D. Plumbley, Fast iterative shrinkage for signal declipping and dequantization, in: iTWIST'18 - International Traveling Workshop on Interactions between low-complexity data models and Sensing Techniques, 2018.

[10] C. Brauer, T. Gerkmann, D. Lorenz, Sparse reconstruction of quantized speech signals, in: 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2016, pp. 5940–5944. doi:10.1109/ICASSP.2016.7472817.

[11] P. Záviška, P. Rajmic, Sparse and cosparse audio dequantization using convex optimization (Mar. 2020). arXiv:2003.04222.
URL https://arxiv.org/abs/1905.00628

[12] P. Rajmic, P. Záviška, V. Veselý, O. Mokrý, A new generalized projection and its application to acceleration of audio declipping, Axioms 8 (3). doi:https://doi.org/10.3390/axioms8030105.

[13] A. Beck, First-Order Methods in Optimization, SIAM-Society for Industrial and Applied Mathematics, 2017.

[14] O. Christensen, Frames and Bases, An Introductory Course, Birkhäuser, Boston, 2008.

[15] O. Christensen, An Introduction to Frames nad Riesz Bases, Birkhäuser, Boston-Basel-Berlin, 2003.

[16] K. Gröchenig, Foundations of time-frequency analysis, Birkhäuser, 2001.

[17] P. Balazs, M. Dörfler, F. Jaillet, N. Holighaus, G. Velasco, Theory, implementation and applications of nonstationary Gabor frames, Journal of computational and applied mathematics 236 (6) (2011) 1481–1496. doi:10.1016/j.cam.2011.09.011.
URL http://www.sciencedirect.com/science/article/pii/S0377042711004900

[18] T. Necciari, P. Balazs, N. Holighaus, P. L. Søndergaard, The erblet transform: An auditory-based time-frequency representation with perfect reconstruction, in: Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on, 2013, pp. 498–502. doi:10.1109/ICASSP.2013.6637697.

[19] R. Gribonval, M. Nikolova, A characterization of proximity operatorsarXiv:1807.04014v2.

[20] W. Etter, Restoration of a discrete-time signal segment by interpolation based on the left-sided and right-sided autoregressive parameters, IEEE Transactions on Signal Processing 44 (5) (1996) 1124–1135. doi:10.1109/78.502326.

[21] J. Lindblom, P. Hedelin, Packet loss concealment based on sinusoidal modeling, in: Speech Coding, 2002, IEEE Workshop Proceedings., IEEE. doi:10.1109/scw.2002.1215725.

[22] C. Rodbro, M. Murthi, S. Andersen, S. Jensen, Hidden markov model-based packet loss concealment for voice over ip, IEEE Transactions on Audio, Speech, and Language Processing 14 (5) (2006) 1609 –1623. doi:10.1109/TSA.2005.858561.

[23] M. Elad, P. Milanfar, R. Rubinstein, Analysis versus synthesis in signal priors, in: Inverse Problems 23 (200), 2005, pp. 947–968.

[24] O. Mokrý, P. Záviška, P. Rajmic, V. Veselý, Introducing SPAIN (SParse Audio INpainter), in: 2019 27th European Signal Processing Conference (EUSIPCO), IEEE, 2019.

[25] A. M. Bruckstein, D. L. Donoho, M. Elad, From sparse solutions of systems of equations to sparse modeling of signals and images, SIAM Review 51 (1) (2009) 34–81.

[26] D. L. Donoho, M. Elad, Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization, Proceedings of The National Academy of Sciences 100 (5) (2003) 2197–2202.

[27] M. Kowalski, K. Siedenburg, M. Dörfler, Social Sparsity! Neighborhood Systems Enrich Structured Shrinkage Operators, Signal Processing, IEEE Transactions on 61 (10) (2013) 2498–2511. doi:10.1109/TSP.2013.2250967.

[28] A. Chambolle, T. Pock, A first-order primal-dual algorithm for convex problems with applications to imaging, Journal of Mathematical Imaging and Vision 40 (1) (2011) 120–145. doi:10.1007/s10851-010-0251-1.

[29] A. Beck, M. Teboulle, A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems, SIAM Journal on Imaging Sciences 2 (1) (2009) 183–202. doi:10.1137/080716542.

[30] K. Siedenburg, M. Doerfler, Persistent time-frequency shrinkage for audio denoising, J. Audio Eng. Soc 61 (1/2) (2013) 29–38.
URL http://www.aes.org/e-lib/browse.cfm?elib=16665

[31] J. J. Moreau, Proximité et dualité dans un espace hilbertien, Bulletin de la société mathématique de France 93 (1965) 273–299.

[32] J.-B. Hiriart-Urruty, C. Lemaréchal, Fundamentals of convex analysis, Springer, 2001.