# Star DGT: a Robust Gabor Transform for Speech Denoising

Vicky Kouni[1*], Holger Rauhut[2] and Theoharis Theoharis[1]

[1*]Department of Informatics and Telecommunications, National and Kapodistrian University of Athens, Athens, Greece.
[2]Chair for Mathematics of Information Processing, RWTH Aachen University, Aachen, Germany.

*Corresponding author(s). E-mail(s): vicky-kouni@di.uoa.gr;
Contributing authors: rauhut@mathc.rwth-aachen.de;
theotheo@di.uoa.gr;

**Abstract**

In this paper, we address the speech denoising problem, where Gaussian and coloured additive noises are to be removed from a given speech signal. Our approach is based on a redundant, analysis-sparse representation of the original speech signal. We pick an eigenvector of the Zauner unitary matrix and – under certain assumptions on the ambient dimension – we use it as window vector to generate a spark deficient Gabor frame. The analysis operator associated with such a frame, is a (highly) redundant Gabor transform, which we use as a sparsifying transform in the denoising procedure. We conduct computational experiments on real-world speech data, using as baseline three Gabor transforms generated by state-of-the-art window vectors in time-frequency analysis and compare their performance to the proposed Gabor transform. The results show that the proposed redundant Gabor transform outperforms previous ones consistently for all types of examined signals of noise.

**Keywords:** Denoising, speech signal, Gabor transform, window vector, spark deficient Gabor frame

# 1 Introduction

Noise is one of the main factors that affect the accuracy of the results in audio processing. Thus audio denoising is one of the most extensively studied inverse problems in signal processing. The task consists in recovering an audio signal $x \in \mathbb{R}^L$ from corrupted linear observations:

$$y = x + e \in \mathbb{R}^L. \tag{1}$$

Noise removal from audio signals is an important first step in applications such as sound classification [1], sound event localization [2], speech recognition [3], dereverberation [4], speech enhancement [5, 6] and source separation [7].

## 1.1 Related Work

In order to address the denoising problem, numerous approaches have emerged, including statistical models [8, 9], empirical mode decomposition [10], spectral subtraction [11, 12], thresholding methods [13, 14], neural networks [15, 16], low-rank models [17, 18], sparse and redundant representations [19–21] and combinations of the aforementioned approaches [22, 23]. Sparse and redundant representations have shown very promising results [24–26], especially when turning to analysis sparsity (also known as co-sparsity) [27–29], which provides flexibility in modelling sparse signals, since it leverages the redundancy of the involved analysis operators. Albeit the authors of [28] are mainly focused on analysis Compressed Sensing [30], they state that using the analysis-prior formulation – with a redundant analysis operator – in denoising is fundamentally different from classical denoising via soft thresholding; we shall call this framework *analysis denoising*. Similarly, [29] explores the superior reconstruction produced by analysis denoising of 1D signals over its synthesis counterpart [31].

## 1.2 Motivation

Our work is inspired by the articles [13, 14, 28, 29, 32, 33]. These publications propose either analysis operators associated with redundant frames (i.e. matrices whose atoms/rows form a frame of the ambient space) with atoms in general position[1], or a finite difference operator (associated with the popular method of total variation [34]), in which many linear dependencies appear for large dimensions. Moreover, in [13, 14, 32], Gabor transforms are combined with thresholding methods for audio denoising. In a similar spirit, we also deploy frames, but we differentiate our approach from related literature in two aspects. First, we choose *spark deficient Gabor frames*, i.e. frames with elements not in general position, so as to leverage the linear dependencies appearing in such frames. Second, our proposed frames are combined with the *analysis denoising formulation*; as mentioned in Section 1.1, analysis denoising can be differentiated from classical thresholding denoising.

[1] equivalently, we call such frames *full spark frames*

The intuition behind employing analysis operators associated with spark deficient frames is based on remarks made in [35]. Therein, the authors demonstrate the advantages of analysis sparsity/cosparsity compared to synthesis sparsity, employing the union-of-subspaces model [36]. According to the latter, sparse signals belong to the union of a combinatorial number of subspaces of specific dimension. Now, it is argued in [35] that the number and dimension of the subspaces (which play a key role in the algorithmic recovery of a signal) associated with the synthesis and analysis sparsity models, can be controlled in terms of the "sparsifying ability"[2] of the involved synthesis and analysis operators, respectively. From a computational perspective – based on the afore-described subspaces' argumentations – analysis sparsity could be more advantageous than its synthesis counterpart, if analysis operators with many linear dependencies among their rows are employed. This is a condition satisfied by spark deficient frames, making them nice candidates for our analysis denoising framework.

Particularly, our choice of spark deficient Gabor frames over other classes of frames that may be spark deficient (e.g. equiangular tight frames [37]) is attributed to the inherent ability of the human auditory system to perform time-frequency analysis [38]. In fact, according to [39], human speech production and speech perception take place in the same area of the human brain, so it is reasonable to assume that speech perception matches speech production to some degree. Hence, since the human auditory system performs Gabor analysis using its own "internal" windows [40], the Gabor transform can be used in practice to resemble the way that the human auditory system works and thus, it is well suited for the application of speech denoising for human listeners. In addition, other proposed transforms for the human auditory system, such as Audlets [41], are also based on the (non-stationary) Gabor transform. To that end, we take advantage of an analysis operator [42], namely star digital Gabor transform (star-DGT), associated with a spark deficient Gabor frame (SDGF). The latter can be generated[3] by time-frequency shifts of any eigenvector of the Zauner unitary matrix [43], under certain assumptions on the signal's dimension. To the best of our knowledge, the efficiency of star-DGT when applied to denoising has not yet been demonstrated. Therefore, it is intriguing to compare the robustness of our proposed Gabor analysis operator to three other Gabor transforms, emerging from state-of-the-art window vectors, by applying all four of them to analysis denoising. Finally, we illustrate the practical importance of our method for real-world speech signals.

[2] that is, the sparsity and cosparsity of a signal with respect to a sparsifying operator

[3] Such a frame can also be generated by the eigenvectors of certain unitaries belonging to the Clifford group, under certain assumptions. However, since the algebraic nature of these assumptions goes beyond the scope of the present paper, we preferred to employ only the Zauner unitary matrix

### 1.3 Key Contributions

The novelty of the proposed method is twofold: (a) we generate a SDGF based on a window vector, associate this SDGF to a highly redundant Gabor analysis operator and use the latter as a sparsifying transform in analysis denoising, (b) we numerically compare the proposed method against three other Gabor analysis operators, based on common windows of time-frequency analysis, on real-world speech data, arguing also about the selection of the lattice parameters. Our experiments show that our method consistently outperforms previous ones, for all speech signals and all types of noise – Gaussian and coloured [44].

### 1.4 Paper organization

The rest of the paper is outlined as follows. In Section 2, we give notation and briefly present the setup for analysis denoising. Section 3 introduces Gabor frames and extends to spark deficient ones, building the desirable SDGF and its associated analysis operator. In Section 4, we describe the experimental settings, while in Section 5, we present two sets of experiments, with corresponding results and evaluation. Lastly, in Section 6 we make some concluding remarks and give potential future directions.

## 2 Gabor denoising setup

### 2.1 Notation

- For a set of indices $N = \{0, 1, \dots, N-1\}$, we write $[N]$.
- (Bra-kets) The set of (column) vectors $|0\rangle, |1\rangle, \dots, |L-1\rangle$ is the standard basis of $\mathbb{C}^L$.
- We write $\mathbb{Z}_L$ for the ring of residues $\bmod L$, that is $\mathbb{Z}_L = \{0 \bmod L, 1 \bmod L, \dots, (L-1) \bmod L\}$.
- For $a, b \in \mathbb{Z}$, we write $a \equiv b (\bmod L)$ for the congruence modulo $L$. Moreover, we write $a \,|\, b$ if $a$ divides $b$; otherwise, we write $a \nmid b$.
- The *power spectral density* (PSD) $S(f)$ of a discrete-time, real-valued signal $z$, is defined as the Fourier transform of the signal's autocorrelation[4] $R_{zz}(l)$, i.e., $S(f) = \sum_{l \in \mathbb{Z}} R_{zz}(l) e^{-ifl}$.
- The support of a signal $x \in \mathbb{R}^L$ is denoted by $\text{supp}(x) = \{i \in [L] : x_i \neq 0\}$. For its cardinality, we write $|\text{supp}(x)|$ and if $|\text{supp}(x)| \leq s << L$, we call $x$ $s$-sparse.
- For $\Phi \in \mathbb{C}^{P \times L}$ $(P > L)$, the quantity $\sigma_s(\Phi x) = \inf\{\|\Phi x - u\|_1 : u \text{ is } s\text{-sparse}\}$ describes the $l_1$-best approximation error to $\Phi x$ by $s$-sparse vectors.

[4] the autocorrelation of a signal is defined as the inner product between the signal and its time-translated version

### 2.2 Analysis denoising formulation

As we mentioned in Section 1, the main idea of speech denoising is to reconstruct a speech signal $x \in \mathbb{R}^L$ from:

$$y = x + e \in \mathbb{R}^L, \tag{2}$$

where $e \in \mathbb{R}^L$, $\|e\|_2 \leq \eta$, corresponds to noise. To do so, we first assume that there exists a redundant sparsifying transform $\Phi \in \mathbb{C}^{P \times L}$ $(P > L)$ called the analysis operator, such that $\Phi x$ is (approximately – in the sense of $\sigma_s(\Phi x)$) sparse. This is an *analysis sparsity model* for $x$.

Using analysis sparsity in denoising, we wish to recover $x$ from $y$. A common approach[5] is the *analysis basis pursuit denoising* problem:

$$\min_{x \in \mathbb{R}^L} \|\Phi x\|_1 \quad \text{subject to} \quad \|x - y\|_2 \leq \eta, \tag{3}$$

or a regularized version[6] [46] of it:

$$\min_{x \in \mathbb{R}^L} \|\Phi x\|_1 + \frac{\mu}{2}\|x - x_0\|_2^2 \quad \text{subject to} \quad \|x - y\|_2 \leq \eta, \tag{4}$$

where $x_0$ denotes an initial guess on $x$ (with standard choices for $x_0$ being $x_0 = A^T y$ or $x_0 = 0$, according to numerical examples of [46]), $\mu > 0$ is a smoothing parameter and $\eta > 0$ an estimate on the noise level.

We will devote the next Section to the construction of a suitable analysis operator $\Phi$.

## 3 Gabor frames

### 3.1 Gabor systems

A *discrete Gabor system* $(g, a, b)$ [47] is defined as a collection of time-frequency shifts of the so-called window vector $g \in \mathbb{C}^L$, expressed as:

$$g_{n,m}(l) = e^{2\pi i m b l / L} g(l - na), \quad l \in [L], \tag{5}$$

where $a$, $b$ denote time and frequency parameters (also known as lattice parameters) respectively, while $n \in [N]$ chosen such that $N = L/a \in \mathbb{N}$ and $m \in [M]$ chosen such that $M = L/b \in \mathbb{N}$ denote time and frequency shift indices respectively. If (5) spans $\mathbb{C}^L$, it is called *Gabor frame* and an equivalent definition of a frame [48] is given below.

[5] since we target at speech denoising, a tractable alternative could be some perceptual variant of the matching pursuit (MP) method, e.g. [45]; however – to our knowledge – perceptual variants of MP account for synthesis sparsity, while we are interested in analysis-sparsity-based denoising

[6] in terms of optimization, it is preferred to solve (4) instead of (3)

**Definition 1** Let $L \in \mathbb{N}$ and $(\phi_p)_{p\in P}$ be a finite subset of $\mathbb{C}^L$. If the inequalities:

$$c_1\|x\|_2^2 \leq \sum_{p\in P} |\langle x, \phi_p\rangle|^2 \leq c_2\|x\|_2^2 \tag{6}$$

hold true for all $x \in \mathbb{C}^L$, for some $0 < c_1 \leq c_2$ (frame bounds), then $(\phi_p)_{p\in P}$ is called a *frame* for $\mathbb{C}^L$.

*Remark 1* The number of elements in $(g, a, b)$ according to (5) is $P = MN = L^2/ab$ and if $(g, a, b)$ is a frame, we have $ab < L$ (the so-called *oversampling* case). A crucial ingredient in order to have a good time-frequency resolution for a signal with respect to a Gabor frame, is the appropriate choice of the time-frequency parameters $a$ and $b$; according to [49], this choice is a challenging task. In the following subsection, we associate two operators with a Gabor frame.

## 3.2 The analysis and synthesis operators associated with a Gabor frame

**Definition 2** Let $\Phi_g : \mathbb{C}^L \mapsto \mathbb{C}^{M\times N}$ denote the *Gabor analysis operator* – also known as DGT[7] – whose action on a signal $x \in \mathbb{C}^L$ is defined as:

$$c(m, n) = \sum_{l=0}^{L-1} x(l)\overline{g(l - na)}e^{-2\pi imbl/L}, \qquad \text{for} \quad m \in [M],\, n \in [N]. \tag{7}$$

**Definition 3** The adjoint of the analysis operator defined in (7), is the *Gabor synthesis operator* $\Phi_g^* : \mathbb{C}^{M\times N} \mapsto \mathbb{C}^L$, whose action on the coefficients $c = c(m, n)$ gives:

$$(\Phi_g^* c)_l = \sum_{n=0}^{N-1}\sum_{m=0}^{M-1} c(m, n)g(l - na)e^{2\pi imbl/L}, \qquad \text{for} \quad l \in [L]. \tag{8}$$

Since we deal with analysis denoising, we will only focus on $\Phi_g$.

## 3.3 Spark deficient Gabor frames

Let us first introduce some basic notions needed in this subsection.

**Definition 4** The symplectic group $\mathrm{SL}(2, \mathbb{Z}_L)$ consists of all matrices:

$$G = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \tag{9}$$

such that $\alpha$, $\beta$, $\gamma$, $\delta \in \mathbb{Z}_L$ and

$$\alpha\delta - \beta\gamma \equiv 1 (\mathrm{mod}\, L). \tag{10}$$

To each such matrix $G$ corresponds (via a projective representation [50]) a unitary matrix $U_G$, given by the explicit formula [51]:

$$U_G = \frac{e^{i\theta}}{\sqrt{L}} \sum_{u,v=0}^{L-1} \tau^{\beta^{-1}(\alpha v^2 - 2uv + \delta u^2)} |u\rangle\langle v|, \tag{11}$$

[7] we will interchangeably use both terms in the sequel

where $\theta$ is an arbitrary phase, $\beta^{-1}$ is the inverse[8] of $\beta \bmod L$ and

$$\tau = -e^{\frac{i\pi}{L}}. \tag{12}$$

**Definition 5** The *spark* of a set $F$ –denoted by $\mathrm{sp}(F)$– of $P$ vectors in $\mathbb{C}^L$ is the size of the smallest linearly dependent subset of $F$. A frame $F$ is full spark if and only if every set of $L$ elements of $F$ is a basis, or equivalently $\mathrm{sp}(F) = L+1$, otherwise it is spark deficient.

Based on the previous definition, a Gabor frame with $P = L^2/ab$ elements of the form (5) is full spark, if and only if every set of $L$ of its elements is a basis. Now, as proven in [48], almost all window vectors generate full spark Gabor frames, so the spark deficient Gabor frames (SDGFs) are generated by exceptional window vectors. Indeed, the following theorem was proven in [51] and informally stated in [50], for the *Zauner* matrix $\mathcal{Z} \in \mathrm{SL}(2, \mathbb{Z}_L)$ given by:

$$\mathcal{Z} = \begin{pmatrix} 0 & -1 \\ 1 & -1 \end{pmatrix} \equiv \begin{pmatrix} 0 & L-1 \\ 1 & L-1 \end{pmatrix}. \tag{13}$$

**Theorem 1** ([51]) *Let $L \in \mathbb{Z}$ such that $2 \nmid L$, $3 \mid L$ and $L$ is square-free. Then, any eigenvector of the Zauner unitary matrix $U_{\mathcal{Z}} \in \mathbb{C}^{L \times L}$ – produced by combining (11) and (13) – generates a spark deficient Gabor frame for $\mathbb{C}^L$.*

According to the previous theorem, the complex-valued eigenvectors of $U_{\mathcal{Z}} \in \mathbb{C}^{L \times L}$ can be used as window vectors to generate – through time-frequency shifts – SDGFs for $\mathbb{C}^L$. Therefore, we can employ Theorem 1 to produce a SDGF and apply its associated analysis operator in (4). To that end, we must first choose an ambient dimension $L$ that fits the assumptions of Theorem 1. Then, we calculate $U_{\mathcal{Z}}$ using (11) and (13) and in the end, perform its spectral decomposition in order to acquire its eigenvectors. Since all the eigenvectors of $U_{\mathcal{Z}}$ generate SDGFs, we may choose an arbitrary one, which we call *star window* from now on and denote it as $g_*$ (see Section 4.2 for a description on how to acquire the desired window vector). We call the analysis operator associated with such a SDGF *star-DGT* and denote it $\Phi_{g_*}$, in order to indicate the dependance on $g_*$. We coin the term "star", due to the slight resemblance of this DGT to a star when plotted in MATLAB, as it is demonstrated in the example of Fig. 1.

*Remark 2* A simple way to choose $L$, is by considering its prime factorization: take $k$ prime numbers $p_1^{\alpha_1}, \ldots, p_k^{\alpha_k}$, with $\alpha_1, \ldots, \alpha_k$ not all a multiple of 2 and $p_1 = 3, p_i \neq 2, i = 2, \ldots, k$, such that $L = 3^{\alpha_1} p_2^{\alpha_2} \cdot \cdots \cdot p_k^{\alpha_k}$. Since $a, b \mid L$, we may also choose $a, b$ to be one, or a multiplication of more than one, prime numbers from the prime

[8] $\beta\beta^{-1} \equiv 1 \bmod L$

factorization of $L$. We have seen empirically that this method for fixing $(L, a, b)$ produces satisfying results, as is illustrated in the sequel.

# 4 Experimental Setup

## 4.1 Signals' description and preprocessing

We run experiments (code available at www.github.com/vicky-k-19/Star-DGT) on 30 real-world, real-valued speech signals, all sampled at 16kHz, taken from *LibriSpeech* corpus [53]. The labels of the signals along with short descriptions can be found in Table 1. The true ambient dimension of each real-world signal does not usually match the conditions of Theorem 1. Hence, we load each signal and use Remark 2 to cut it down to a specific ambient dimension (which will be referred to in the sequel as *artificial dimension*) $L$, being as close as possible to its true dimension, in order to both denoise a meaningful part of the signal and meet the conditions of Theorem 1.

## 4.2 Experimental settings

1. We examine different pairs of time-frequency parameters $(a, b)$, deploying Remark 2.
2. We use the power iteration method [55], which yields the largest in magnitude eigenvalue, with corresponding eigenvector, of $U_{\mathcal{Z}}$. We keep the real part of this complex-valued eigenvector and set it as our desired star window vector $g_*$.

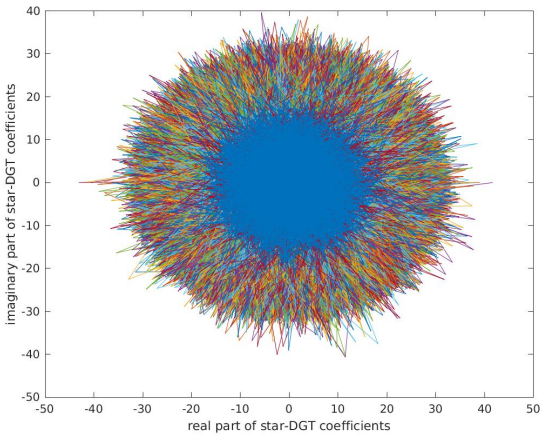



**Fig. 1**: Star-DGT coefficients $c(m, n)$ in matrix form (plotted real vs imaginary part – different colors correspond to different entries of the matrix), for the audio signal *Glockenspiel* [52] (262144 samples, $(L, a, b) = (200583, 19, 23)$)

**Table 1**: Details of example speech signals chosen from Librispeech corpus

| # | Label | True ambient dimension | Artificial dimension $L$ | Types of noise added |
|---|---|---|---|---|
| 1 | 251-136532-0014 | 36240 | 33915 | Gaussian and pink |
| 2 | 8842-304647-0007 | 27680 | 27531 | Gaussian and blue |
| 3 | 2035-147960-0013 | 42800 | 41769 | Gaussian and pink |
| 4 | 1462-170145-0020 | 34400 | 33915 | Gaussian and blue |
| 5 | 6241-61943-0002 | 43760 | 43605 | Gaussian and blue |
| 6 | 5338-284437-0025 | 31040 | 29835 | Gaussian and blue |
| 7 | 3752-4944-0042 | 51360 | 51051 | Gaussian and blue |
| 8 | 5694-64038-0013 | 52880 | 51051 | Gaussian and pink |
| 9 | 5895-34615-0001 | 52880 | 51051 | Gaussian and pink |
| 10 | 2428-83699-0035 | 43600 | 41769 | Gaussian and pink |
| 11 | 2803-154320-0006 | 34880 | 33915 | Gaussian and blue |
| 12 | 3752-4944-0008 | 31040 | 29835 | Gaussian and pink |
| 13 | 1272-135031-0014 | 27840 | 24225 | Gaussian and pink |
| 14 | 1272-141231-0016 | 29600 | 29325 | Gaussian and blue |
| 15 | 7850-281318-0020 | 34799 | 31395 | Gaussian and pink |
| 16 | 2035-147960-0015 | 24960 | 24225 | Gaussian and blue |
| 17 | 2035-147961-0017 | 32160 | 31395 | Gaussian and blue |
| 18 | 2035-147961-0027 | 46000 | 43263 | Gaussian and pink |
| 19 | 6295-244435-0024 | 67120 | 65379 | Gaussian and blue |
| 20 | 3576-138058-0003 | 40160 | 38709 | Gaussian and blue |
| 21 | 5338-284437-0019 | 49840 | 49725 | Gaussian and blue |
| 22 | 2412-153948-0015 | 55440 | 51129 | Gaussian and pink |
| 23 | 2803-154320-0009 | 44240 | 43263 | Gaussian and blue |
| 24 | 5694-64038-0007 | 37920 | 34017 | Gaussian and pink |
| 25 | 5694-64038-0016 | 41360 | 38019 | Gaussian and blue |
| 26 | 5895-34622-0018 | 79680 | 78039 | Gaussian and blue |
| 27 | 5338-284437-0014 | 29520 | 26013 | Gaussian and blue |
| 28 | 6241-61943-0023 | 69680 | 66033 | Gaussian and pink |
| 29 | 6319-57405-0011 | 59360 | 57753 | Gaussian and pink |
| 30 | 6295-244435-0003 | 51920 | 50025 | Gaussian and blue |

3. We construct – using the MATLAB package *LTFAT* [52] – four different Gabor frames with their associated analysis operators/DGTs: $\Phi_{g_1}$, $\Phi_{g_2}$, $\Phi_{g_3}$ and $\Phi_{g_*}$, corresponding to a Gaussian, a Hann, a Hamming and the star window vector, respectively. Since we process real-valued signals, we alter the four analysis operators to compute only the DGT coefficients of positive frequencies instead of the full DGT coefficients.
4. We consider Gaussian and coloured additive noises; the Gaussian noise has zero mean and the coloured noises are chosen to be pink and blue. Pink noise [56] is a signal with a power spectral density (see Section 2.1 for a formal definition) $S(f)$ inversely proportional to the frequency $f$ of the signal. Blue noise [57] is a signal with a power spectral density $S(f)$ proportional to the frequency $f$ of the signal. Formally speaking, it holds $k_1/f^\gamma \leq S(f) \leq k_2/f^\gamma$, where $\gamma = 1$ for pink noise and $\gamma = -1$ for blue noise.
5. For the Gaussian noise, we employ the MATLAB built-in function LINSPACE to generate a vector $\sigma_G$ of 100 evenly spaced points in the interval $[0.001, 0.01]$. This vector is used as a scaling factor controlling the standard deviation of the Gaussian noise. In a similar manner for the coloured noises[9], we employ LINSPACE to generate a vector $\sigma_{col}$ of 100 evenly spaced points in the interval $[0.001, 0.01]$. Each value of $\sigma_{col}$ serves

[9] in the rest of the paper, when we speak of coloured noises, we mean the examined cases of pink and blue noise

**Table 2**: Mean scores of MSE and output SNR (as defined in [54]) achieved by all four DGTs, averaged over all 30 examined speech signals, for fixed values of $\sigma_G$ and $\sigma_{col}$. Bold letters indicate the best mean scores among all DGTs.

| Windows \ Noises | MSE | | | Output SNR | | |
|---|---|---|---|---|---|---|
| | Gaussian | Blue | Pink | Gaussian | Blue | Pink |
| star | $\mathbf{6.4956 \cdot 10^{-4}}$ | $\mathbf{8.3 \cdot 10^{-4}}$ | $\mathbf{6.5035 \cdot 10^{-4}}$ | **8.6582** | **7.59345** | **8.65298** |
| Gaussian | $9.1771 \cdot 10^{-4}$ | $1.2264 \cdot 10^{-3}$ | $8.127 \cdot 10^{-4}$ | 7.1573 | 5.8980 | 7.6851 |
| Hann | $9.1768 \cdot 10^{-4}$ | $1.2264 \cdot 10^{-3}$ | $8.1267 \cdot 10^{-4}$ | 7.1575 | 5.8983 | 7.6853 |
| Hamming | $9.1769 \cdot 10^{-4}$ | $1.2265 \cdot 10^{-3}$ | $8.1279 \cdot 10^{-4}$ | 7.1574 | 5.8978 | 7.6846 |

(a) Fixed $\sigma_G = \sigma_{col} = 0.01$ for the scaling factor controlling the std of the Gaussian noise and the amplitude of the coloured noises

| Windows \ Noises | MSE | | | Output SNR | | |
|---|---|---|---|---|---|---|
| | Gaussian | Blue | Pink | Gaussian | Blue | Pink |
| star | $\mathbf{3.1934 \cdot 10^{-4}}$ | $\mathbf{4.737 \cdot 10^{-4}}$ | $\mathbf{4.4463 \cdot 10^{-4}}$ | **11.7418** | **10.0294** | **10.3044** |
| Gaussian | $5.8417 \cdot 10^{-4}$ | $8.3276 \cdot 10^{-4}$ | $6.2148 \cdot 10^{-4}$ | 9.119 | 7.5792 | 8.8501 |
| Hann | $5.8418 \cdot 10^{-4}$ | $8.3276 \cdot 10^{-4}$ | $6.2150 \cdot 10^{-4}$ | 9.1189 | 7.5792 | 8.85 |
| Hamming | $5.8423 \cdot 10^{-4}$ | $8.3276 \cdot 10^{-4}$ | $6.2152 \cdot 10^{-4}$ | 9.1185 | 7.5792 | 8.8499 |

(b) Fixed $\sigma_G = \sigma_{col} = 0.001$ for the scaling factor controlling the std of the Gaussian noise and the amplitude of the coloured noises

**Table 3**: Example speech signal (**251-136532-0014**), contaminated by additive Gaussian and pink noises

| Windows \ $(a,b)$ | MSEs | | | | |
|---|---|---|---|---|---|
| | $(15,15)$ | $(5,17)$ | $(7,19)$ | $(17,19)$ | $(21,21)$ |
| Gaussian | $8.4929 \cdot 10^{-4}$ | $8.4846 \cdot 10^{-4}$ | $8.4233 \cdot 10^{-4}$ | $8.4488 \cdot 10^{-4}$ | $8.4188 \cdot 10^{-4}$ |
| Hann | $8.4917 \cdot 10^{-4}$ | $8.4858 \cdot 10^{-4}$ | $8.4222 \cdot 10^{-4}$ | $8.4502 \cdot 10^{-4}$ | $8.4196 \cdot 10^{-4}$ |
| Hamming | $8.4936 \cdot 10^{-4}$ | $8.4830 \cdot 10^{-4}$ | $8.4234 \cdot 10^{-4}$ | $8.4490 \cdot 10^{-4}$ | $8.4199 \cdot 10^{-4}$ |
| star | $\mathbf{7.4783 \cdot 10^{-4}}$ | $\mathbf{8.1694 \cdot 10^{-4}}$ | $\mathbf{7.5914 \cdot 10^{-4}}$ | $6.7464 \cdot 10^{-4}$ | $\mathbf{7.0497 \cdot 10^{-4}}$ |

(a) Additive Gaussian noise, with standard deviation $\sigma_G = 0.001$

| Windows \ $(a,b)$ | MSEs | | | | |
|---|---|---|---|---|---|
| | $(15,15)$ | $(5,17)$ | $(7,19)$ | $(17,19)$ | $(21,21)$ |
| Gaussian | $5.3213 \cdot 10^{-4}$ | $5.3660 \cdot 10^{-4}$ | $5.3533 \cdot 10^{-4}$ | $5.3625 \cdot 10^{-4}$ | $5.4130 \cdot 10^{-4}$ |
| Hann | $5.3217 \cdot 10^{-4}$ | $5.3655 \cdot 10^{-4}$ | $5.3521 \cdot 10^{-4}$ | $5.3620 \cdot 10^{-4}$ | $5.4122 \cdot 10^{-4}$ |
| Hamming | $5.3213 \cdot 10^{-4}$ | $5.3655 \cdot 10^{-4}$ | $5.3528 \cdot 10^{-4}$ | $5.3632 \cdot 10^{-4}$ | $5.4111 \cdot 10^{-4}$ |
| star | $\mathbf{4.5283 \cdot 10^{-4}}$ | $\mathbf{4.7100 \cdot 10^{-4}}$ | $\mathbf{4.6333 \cdot 10^{-4}}$ | $3.5476 \cdot 10^{-4}$ | $\mathbf{4.2386 \cdot 10^{-4}}$ |

(b) Additive pink noise, with amplitude's scaling factor $\sigma_{col} = 0.001$

as a scaling factor controlling the amplitude of the coloured noises. Note that when $\sigma_{col}$ reaches its top value 0.01, the amplitude of the coloured noises is almost equal to the amplitude of the examined signals. For each value of $\sigma_k \in [0.001, 0.01]$, $k = G$, $col$, we follow the procedure described below:

- we take noisy measurements $y = x + e$, where $e$ denotes either Gaussian or coloured noise. For the estimate on the noise level $\eta$, we set $\eta = \|y -$

**Table 4**: Example speech signal (**2035-147960-0013**), contaminated by additive Gaussian and pink noises

| | MSEs | | | | |
|---|---|---|---|---|---|
| Windows \ $(a,b)$ | $(9,9)$ | $(7,13)$ | $(9,17)$ | $(13,17)$ | $(17,17)$ |
| Gaussian | $9.1225 \cdot 10^{-4}$ | $9.1113 \cdot 10^{-4}$ | $9.1134 \cdot 10^{-4}$ | $9.0846 \cdot 10^{-4}$ | $9.1484 \cdot 10^{-4}$ |
| Hann | $9.1230 \cdot 10^{-4}$ | $9.1111 \cdot 10^{-4}$ | $9.1126 \cdot 10^{-4}$ | $9.0833 \cdot 10^{-4}$ | $9.1492 \cdot 10^{-4}$ |
| Hamming | $9.1221 \cdot 10^{-4}$ | $9.1108 \cdot 10^{-4}$ | $9.1130 \cdot 10^{-4}$ | $9.0838 \cdot 10^{-4}$ | $9.1492 \cdot 10^{-4}$ |
| star | $\mathbf{8.5635 \cdot 10^{-4}}$ | $\mathbf{8.0955 \cdot 10^{-4}}$ | $\mathbf{8.2828 \cdot 10^{-4}}$ | *7.7499 · 10⁻⁴* | $\mathbf{8.1464 \cdot 10^{-4}}$ |

(a) Additive Gaussian noise, with standard deviation $\sigma_G = 0.001$

| | MSEs | | | | |
|---|---|---|---|---|---|
| Windows \ $(a,b)$ | $(9,9)$ | $(7,13)$ | $(9,17)$ | $(13,17)$ | $(17,17)$ |
| Gaussian | $5.8699 \cdot 10^{-4}$ | $5.8925 \cdot 10^{-4}$ | $5.8942 \cdot 10^{-4}$ | $5.9071 \cdot 10^{-4}$ | $5.8948 \cdot 10^{-4}$ |
| Hann | $5.8710 \cdot 10^{-4}$ | $5.8921 \cdot 10^{-4}$ | $5.8945 \cdot 10^{-4}$ | $5.9080 \cdot 10^{-4}$ | $5.8949 \cdot 10^{-4}$ |
| Hamming | $5.8703 \cdot 10^{-4}$ | $5.8928 \cdot 10^{-4}$ | $5.8961 \cdot 10^{-4}$ | $5.9060 \cdot 10^{-4}$ | $5.8943 \cdot 10^{-4}$ |
| star | $\mathbf{5.3307 \cdot 10^{-4}}$ | $\mathbf{5.3072 \cdot 10^{-4}}$ | $\mathbf{4.9653 \cdot 10^{-4}}$ | *4.8565 · 10⁻⁴* | $\mathbf{4.8874 \cdot 10^{-4}}$ |

(b) Additive pink noise, with amplitude's scaling factor $\sigma_{col} = 0.001$

**Table 5**: Example speech signal (**5694-64038-0013**), contaminated by additive Gaussian and pink noises

| | MSEs | | | | |
|---|---|---|---|---|---|
| Windows \ $(a,b)$ | $(11,11)$ | $(13,13)$ | $(13,21)$ | $(11,17)$ | $(13,17)$ |
| Gaussian | $8.7630 \cdot 10^{-4}$ | $8.7634 \cdot 10^{-4}$ | $8.7914 \cdot 10^{-4}$ | $8.7299 \cdot 10^{-4}$ | $8.7429 \cdot 10^{-4}$ |
| Hann | $8.7643 \cdot 10^{-4}$ | $8.7641 \cdot 10^{-4}$ | $8.7918 \cdot 10^{-4}$ | $8.7300 \cdot 10^{-4}$ | $8.7419 \cdot 10^{-4}$ |
| Hamming | $8.7629 \cdot 10^{-4}$ | $8.7639 \cdot 10^{-4}$ | $8.7909 \cdot 10^{-4}$ | $8.7301 \cdot 10^{-4}$ | $8.7431 \cdot 10^{-4}$ |
| star | $\mathbf{8.2926 \cdot 10^{-4}}$ | $\mathbf{8.0809 \cdot 10^{-4}}$ | $\mathbf{7.6079 \cdot 10^{-4}}$ | $\mathbf{7.5266 \cdot 10^{-4}}$ | *7.5009 · 10⁻⁴* |

(a) Additive Gaussian noise, with standard deviation $\sigma_G = 0.001$

| | MSEs | | | | |
|---|---|---|---|---|---|
| Windows \ $(a,b)$ | $(11,11)$ | $(13,13)$ | $(13,21)$ | $(11,17)$ | $(13,17)$ |
| Gaussian | $5.5950 \cdot 10^{-4}$ | $5.5945 \cdot 10^{-4}$ | $5.6133 \cdot 10^{-4}$ | $5.5808 \cdot 10^{-4}$ | $5.5888 \cdot 10^{-4}$ |
| Hann | $5.5949 \cdot 10^{-4}$ | $5.5943 \cdot 10^{-4}$ | $5.6131 \cdot 10^{-4}$ | $5.5812 \cdot 10^{-4}$ | $5.5890 \cdot 10^{-4}$ |
| Hamming | $5.5949 \cdot 10^{-4}$ | $5.5958 \cdot 10^{-4}$ | $5.6128 \cdot 10^{-4}$ | $5.5800 \cdot 10^{-4}$ | $5.5897 \cdot 10^{-4}$ |
| star | $\mathbf{5.1971 \cdot 10^{-4}}$ | $\mathbf{4.9725 \cdot 10^{-4}}$ | $\mathbf{4.4889 \cdot 10^{-4}}$ | $\mathbf{4.7244 \cdot 10^{-4}}$ | *4.3773 · 10⁻⁴* |

(b) Additive pink noise, with amplitude's scaling factor $\sigma_{col} = 0.001$

$x\|_2$. This is considered a standard scenario in the numerical examples of [46].

- we solve – using the MATLAB package *TFOCS* [46] – four different instances of (4), one for each of the four DGTs. For TFOCS, we set $x_0 = 0$, $z_0 = [\,]$; for each of the instances $i = 1, 2, 3, *$, we set the smoothing parameter $\mu_i = 10^{-1}\|\Phi_{g_i} x\|_\infty$, since we noticed an improved performance of the solving algorithm when $\mu$ is a function of $\Phi_{g_i}$ (we chose the scaling factor $10^{-1}$ and the $l_\infty$-norm after some empirical tests) and employ the *solver_BPDN_W* solver.
- from the solving procedure, we obtain four different estimators for $x$, namely $\hat{x}_1$, $\hat{x}_2$, $\hat{x}_3$, $\hat{x}_*$ and their corresponding mean-squared errors

**Table 6**: Example speech signal (**5338-284437-0025**), contaminated by additive Gaussian and blue noises

| | MSEs | | | | |
|---|---|---|---|---|---|
| Windows \ $(a,b)$ | $(9,9)$ | $(15,15)$ | $(5,13)$ | $(5,17)$ | $(13,17)$ |
| Gaussian | $12 \cdot 10^{-4}$ | $12 \cdot 10^{-4}$ | $12 \cdot 10^{-4}$ | $12 \cdot 10^{-4}$ | $12 \cdot 10^{-4}$ |
| Hann | $12 \cdot 10^{-4}$ | $12 \cdot 10^{-4}$ | $12 \cdot 10^{-4}$ | $12 \cdot 10^{-4}$ | $12 \cdot 10^{-4}$ |
| Hamming | $12 \cdot 10^{-4}$ | $12 \cdot 10^{-4}$ | $12 \cdot 10^{-4}$ | $12 \cdot 10^{-4}$ | $12 \cdot 10^{-4}$ |
| star | $\mathbf{11 \cdot 10^{-4}}$ | $\mathbf{9.8065 \cdot 10^{-4}}$ | $\mathbf{11 \cdot 10^{-4}}$ | $\mathbf{11 \cdot 10^{-4}}$ | $9.7897 \cdot 10^{-4}$ |

(a) Additive Gaussian noise, with standard deviation $\sigma_G = 0.001$

| | MSEs | | | | |
|---|---|---|---|---|---|
| Windows \ $(a,b)$ | $(9,9)$ | $(15,15)$ | $(5,13)$ | $(5,17)$ | $(13,17)$ |
| Gaussian | $7.8626 \cdot 10^{-4}$ | $7.8715 \cdot 10^{-4}$ | $7.8627 \cdot 10^{-4}$ | $7.8690 \cdot 10^{-4}$ | $7.8674 \cdot 10^{-4}$ |
| Hann | $7.8629 \cdot 10^{-4}$ | $7.8689 \cdot 10^{-4}$ | $7.8650 \cdot 10^{-4}$ | $7.8689 \cdot 10^{-4}$ | $7.8696 \cdot 10^{-4}$ |
| Hamming | $7.8627 \cdot 10^{-4}$ | $7.8745 \cdot 10^{-4}$ | $7.8623 \cdot 10^{-4}$ | $7.8676 \cdot 10^{-4}$ | $7.8677 \cdot 10^{-4}$ |
| star | $\mathbf{6.8879 \cdot 10^{-4}}$ | $\mathbf{6.6695 \cdot 10^{-4}}$ | $\mathbf{7.0199 \cdot 10^{-4}}$ | $\mathbf{6.6884 \cdot 10^{-4}}$ | $6.3867 \cdot 10^{-4}$ |

(b) Additive blue noise, with amplitude's scaling factor $\sigma_{col} = 0.001$

**Table 7**: Example speech signal (**6241-61943-0002**), contaminated by additive Gaussian and blue noises

| | MSEs | | | | |
|---|---|---|---|---|---|
| Windows \ $(a,b)$ | $(9,9)$ | $(5,17)$ | $(9,17)$ | $(17,19)$ | $(19,19)$ |
| Gaussian | $5.0929 \cdot 10^{-4}$ | $5.1122 \cdot 10^{-4}$ | $5.1164 \cdot 10^{-4}$ | $5.0961 \cdot 10^{-4}$ | $5.1202 \cdot 10^{-4}$ |
| Hann | $5.0929 \cdot 10^{-4}$ | $5.1119 \cdot 10^{-4}$ | $5.1166 \cdot 10^{-4}$ | $5.0961 \cdot 10^{-4}$ | $5.1202 \cdot 10^{-4}$ |
| Hamming | $5.0928 \cdot 10^{-4}$ | $5.1121 \cdot 10^{-4}$ | $5.1160 \cdot 10^{-4}$ | $5.0957 \cdot 10^{-4}$ | $5.1201 \cdot 10^{-4}$ |
| star | $\mathbf{4.9231 \cdot 10^{-4}}$ | $\mathbf{4.6263 \cdot 10^{-4}}$ | $\mathbf{4.7426 \cdot 10^{-4}}$ | $4.5764 \cdot 10^{-4}$ | $\mathbf{4.6173 \cdot 10^{-4}}$ |

(a) Additive Gaussian noise, with standard deviation $\sigma_G = 0.001$

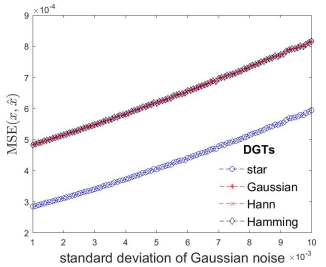
| | MSEs | | | | |
|---|---|---|---|---|---|
| Windows \ $(a,b)$ | $(9,9)$ | $(5,17)$ | $(9,17)$ | $(17,19)$ | $(19,19)$ |
| Gaussian | $2.6371 \cdot 10^{-4}$ | $2.6358 \cdot 10^{-4}$ | $2.6336 \cdot 10^{-4}$ | $2.6356 \cdot 10^{-4}$ | $2.6356 \cdot 10^{-4}$ |
| Hann | $2.6373 \cdot 10^{-4}$ | $2.6360 \cdot 10^{-4}$ | $2.6338 \cdot 10^{-4}$ | $2.6359 \cdot 10^{-4}$ | $2.6359 \cdot 10^{-4}$ |
| Hamming | $2.6371 \cdot 10^{-4}$ | $2.6359 \cdot 10^{-4}$ | $2.6340 \cdot 10^{-4}$ | $2.6359 \cdot 10^{-4}$ | $2.6359 \cdot 10^{-4}$ |
| star | $\mathbf{2.3890 \cdot 10^{-4}}$ | $\mathbf{2.3582 \cdot 10^{-4}}$ | $\mathbf{2.2962 \cdot 10^{-4}}$ | $1.9016 \cdot 10^{-4}$ | $\mathbf{2.2279 \cdot 10^{-4}}$ |

(b) Additive blue noise, with amplitude's scaling factor $\sigma_{col} = 0.001$

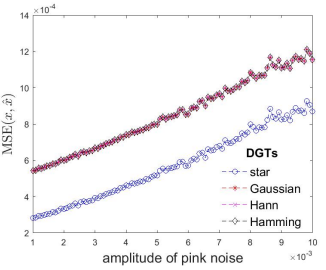
(MSEs):

$$\mathrm{MSE}(x, \hat{x}_i) = \frac{1}{L} \sum_{l=1}^{L} (x(l) - \hat{x}_i(l))^2, \qquad i = 1, 2, 3, *. \tag{14}$$

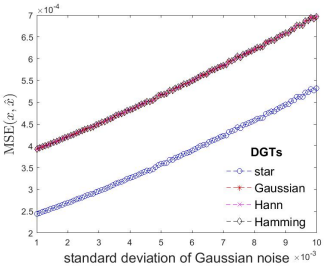
We choose MSE as a reconstruction quality measure, since it represents the standard choice for denoising problems [20], [24], [25], [29], [58].



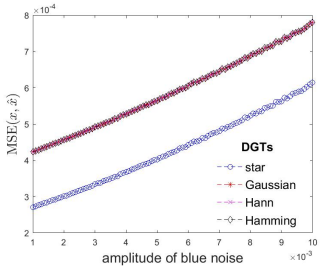

(a) 251-136532-0014, $(L, a, b) = (33915, 51, 19)$, Gaussian noise



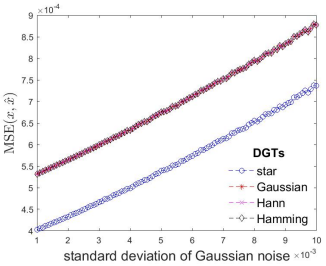

(b) 251-136532-0014, $(L, a, b) = (33915, 51, 19)$, pink noise



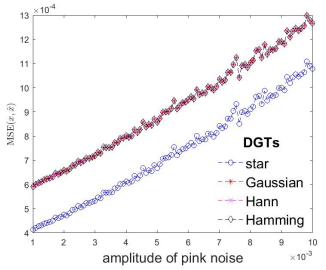

(c) 8842-304647-0007, $(L, a, b) = (27531, 19, 23)$, Gaussian noise



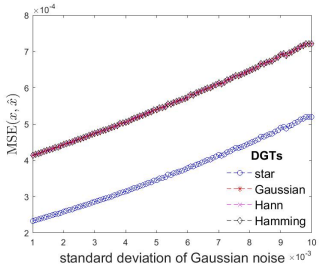

(d) 8842-304647-0007, $(L, a, b) = (27531, 19, 23)$, blue noise



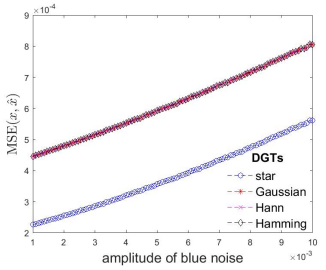

(e) 2035-147960-0013, $(L, a, b) = (41769, 21, 17)$, Gaussian noise




(f) 2035-147960-0013, $(L, a, b) = (41769, 21, 17)$, pink noise

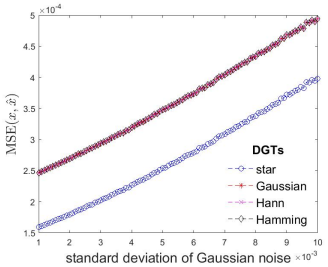



(g) 1462-170145-0020, $(L, a, b) = (33915, 51, 19)$, Gaussian noise

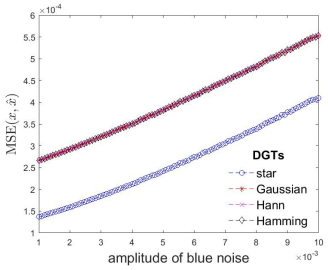



(h) 1462-170145-0020, $(L, a, b) = (33915, 51, 19)$, blue noise

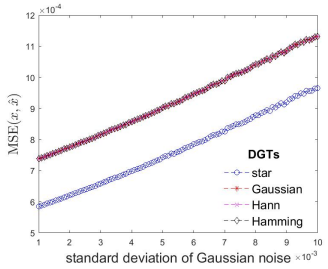
**Fig. 2**: Rate of denoising success for 4 speech signals with different parameters $(L, a, b)$, contaminated by Gaussian (left) and coloured (right) noise. Note that 3 of the 4 methods roughly coincide.



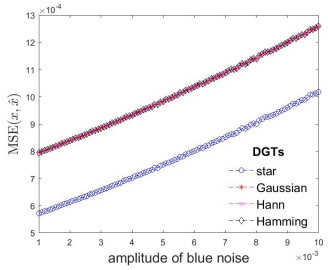

(a) 6241-61943-0002, $(L, a, b) = (43605, 51, 19)$, Gaussian noise



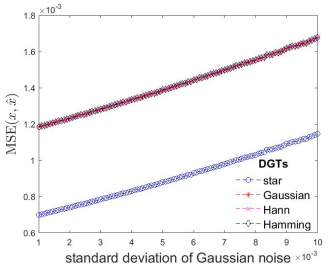

(b) 6241-61943-0002, $(L, a, b) = (43605, 51, 19)$, pink noise



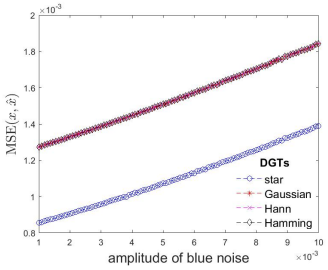

(c) 5338-284437-0025, $(L, a, b) = (29835, 17, 13)$, Gaussian noise



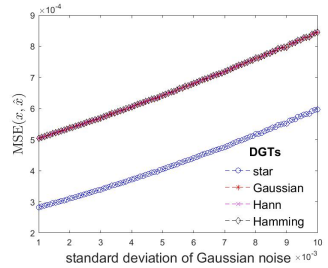

(d) 5338-284437-0025, $(L, a, b) = (29835, 17, 13)$, blue noise



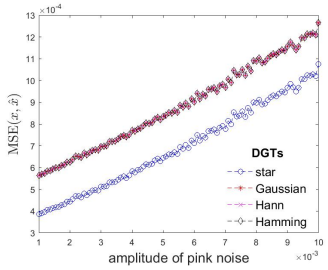

(e) 3752-4944-0042, $(L, a, b) = (51051, 33, 17)$, Gaussian noise




(f) 3752-4944-0042, $(L, a, b) = (51051, 33, 17)$, blue noise

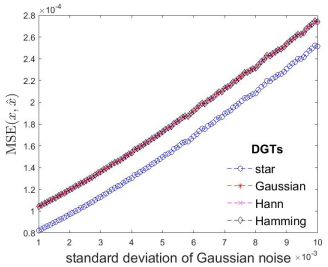



(g) 5694-64038-0013, $(L, a, b) = (51051, 33, 17)$, Gaussian noise

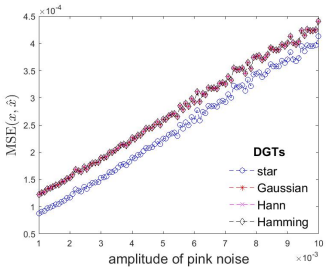



(h) 5694-64038-0013, $(L, a, b) = (51051, 33, 17)$, pink noise

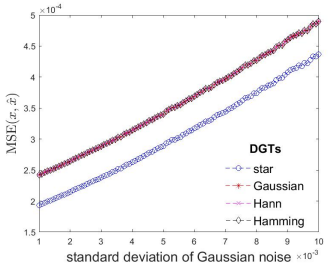
**Fig. 3**: Rate of denoising success for 4 speech signals with different parameters $(L, a, b)$, contaminated by Gaussian (left) and coloured (right) noise. Note that 3 of the 4 methods roughly coincide.



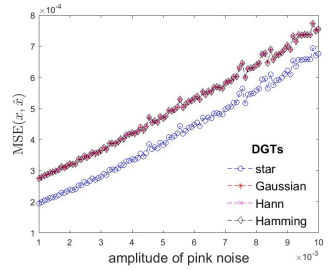

(a) 5895-34615-0001, $(L, a, b) = (51051, 33, 17)$, Gaussian noise



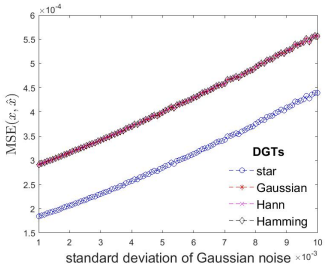

(b) 5895-34615-0001, $(L, a, b) = (51051, 33, 17)$, pink noise



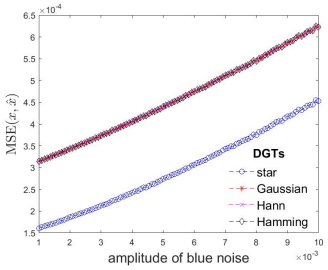

(c) 2428-83699-0035, $(L, a, b) = (41769, 21, 17)$, Gaussian noise



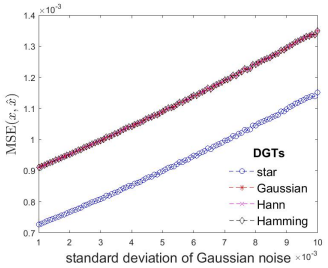

(d) 2428-83699-0035, $(L, a, b) = (41769, 21, 17)$, pink noise



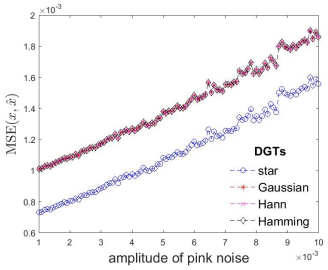

(e) 2803-154320-0006, $(L, a, b) = (33915, 51, 19)$, Gaussian noise




(f) 2803-154320-0006, $(L, a, b) = (33915, 51, 19)$, blue noise




(g) 3752-4944-0008, $(L, a, b) = (29835, 17, 13)$, Gaussian noise




(h) 3752-4944-0008, $(L, a, b) = (29835, 17, 13)$, pink noise

**Fig. 4**: Rate of denoising success for 4 speech signals with different parameters $(L, a, b)$, contaminated by Gaussian (left) and coloured (right) noise. Note that 3 of the 4 methods roughly coincide.

**Table 8**: Example speech signal (**8842-304647-0007**), contaminated by additive Gaussian and blue noises

| Windows \ $(a,b)$ | MSEs | | | | |
|---|---|---|---|---|---|
| | $(9,9)$ | $(9,23)$ | $(7,19)$ | $(19,19)$ | $(19,23)$ |
| Gaussian | $7.2315 \cdot 10^{-4}$ | $7.2621 \cdot 10^{-4}$ | $7.1806 \cdot 10^{-4}$ | $7.2532 \cdot 10^{-4}$ | $7.1924 \cdot 10^{-4}$ |
| Hann | $7.2314 \cdot 10^{-4}$ | $7.2605 \cdot 10^{-4}$ | $7.1806 \cdot 10^{-4}$ | $7.2535 \cdot 10^{-4}$ | $7.1915 \cdot 10^{-4}$ |
| Hamming | $7.2321 \cdot 10^{-4}$ | $7.2609 \cdot 10^{-4}$ | $7.1803 \cdot 10^{-4}$ | $7.2519 \cdot 10^{-4}$ | $7.1919 \cdot 10^{-4}$ |
| star | $\mathbf{6.7449 \cdot 10^{-4}}$ | $\mathbf{6.1704 \cdot 10^{-4}}$ | $\mathbf{6.0620 \cdot 10^{-4}}$ | $\mathbf{6.2294 \cdot 10^{-4}}$ | $5.4646 \cdot 10^{-4}$ |

(a) Additive Gaussian noise, with standard deviation $\sigma_G = 0.001$

| Windows \ $(a,b)$ | MSEs | | | | |
|---|---|---|---|---|---|
| | $(9,9)$ | $(9,23)$ | $(7,19)$ | $(19,19)$ | $(19,23)$ |
| Gaussian | $4.1860 \cdot 10^{-4}$ | $4.1906 \cdot 10^{-4}$ | $4.1881 \cdot 10^{-4}$ | $4.1925 \cdot 10^{-4}$ | $4.1887 \cdot 10^{-4}$ |
| Hann | $4.1863 \cdot 10^{-4}$ | $4.1903 \cdot 10^{-4}$ | $4.1879 \cdot 10^{-4}$ | $4.1935 \cdot 10^{-4}$ | $4.1887 \cdot 10^{-4}$ |
| Hamming | $4.1864 \cdot 10^{-4}$ | $4.1899 \cdot 10^{-4}$ | $4.1885 \cdot 10^{-4}$ | $4.1932 \cdot 10^{-4}$ | $4.1884 \cdot 10^{-4}$ |
| star | $\mathbf{3.6969 \cdot 10^{-4}}$ | $\mathbf{2.8825 \cdot 10^{-4}}$ | $\mathbf{3.3767 \cdot 10^{-4}}$ | $\mathbf{3.2135 \cdot 10^{-4}}$ | $2.8677 \cdot 10^{-4}$ |

(b) Additive blue noise, with amplitude's scaling factor $\sigma_{col} = 0.001$

# 5 Experiments and Results

We randomly choose 12 of the 30 signals to run the experiments of Sections 5.1 and 5.2. For all 30 signals, we gather the mean scores of the experiments of Sections 5.1 and 5.2 in Table 2. At this point, we would like to note that there is sufficiently strong interest in the research area of sparsity-based denoising, making comparisons with other model-based methods unnecessary (cf. Section 1.1). Data-driven approaches could offer a performance advantage, but they assume the existence of a large training dataset with ground truths, which is not available in all cases; thus, model-based methods retain their value.

## 5.1 Fixed $(L, a, b)$ with varying $\sigma_G$ and $\sigma_{col}$

We fix for each of the 12 signals the lattice parameters $a, b$, with respect to each artificial dimension $L$. We add to all signals zero-mean Gaussian noise with varying standard deviation $\sigma_G$ and perform analysis denoising for each value of $\sigma_G$, as explained in Section 4.2. For the coloured noise cases, we randomly split the set of 12 signals into two subsets, of 6 signals each. We add to the signals of the first and second subset blue and pink noise, respectively, with varying amplitude controlled by $\sigma_{col}$, and perform analysis denoising for each value of $\sigma_{col}$, as described in Section 4.2. The left column in Figs. 2-4 demonstrates for different signals, how the 4 resulting MSEs scale in the case of the Gaussian noise, as its standard deviation increases. Clearly, our proposed DGT outperforms the rest of DGTs, consistently for all signals and for different choices of artificial dimension with time-frequency parameters. Similarly, the right column in Figs. 2-4 demonstrates for different signals, how the 4 resulting MSEs scale in the case of blue and pink noise, as the scaling factor of each coloured noise's amplitude increases. We observe that star-DGT is more robust than the rest of the DGTs, even when the amplitude of each

coloured noise is almost equal to the amplitude of the speech signal to which it is added.

### 5.2 Fixed $L$, $\sigma_G$, $\sigma_{col}$, with varying $(a, b)$

We randomly pick 6 out of the 12 speech signals (preferring to examine signals with different artificial dimensions). For each signal, we alter the time-frequency parameters $a, b$ with respect to its artificial dimension. We consider fixed values of $\sigma_G$ and $\sigma_{col}$, i.e. $\sigma_G = \sigma_{col} = 0.001$, serving as the standard deviation of the Gaussian noise and the scaling factor controlling the coloured noises' amplitude, respectively. For different pairs of $(a, b)$, we add zero-mean Gaussian noise to all six signals, blue noise to three of the six signals and pink noise to the rest of them. Finally, we denoise all signals for all types of noise and present the resulting MSEs in Tables 3-8. For all choices of $(a, b)$, star-DGT (indicated in bold in each subtable) outperforms the baseline DGTs, consistently for all signals, for both Gaussian and coloured noises. Additionally, we see that among all examined pairs of lattice parameters, star-DGT achieves the smallest MSE (indicated in magenta italics in each subtable) when $a, b$ are chosen as the two largest factors in the prime factorization of $L$; the rest of the DGTs do not seem to benefit much from this selection. On the other hand, among all examined choices of $(a, b)$, star-DGT performs slightly worse when $a = b$. For example, as indicated in Tables 5 and 8, star-DGT reaches a slightly bigger MSE when $a = b = 11$ and $a = b = 3^2$, respectively. The afore-described experimental result triggers us to formulate it as the following (informal) mathematical hypothesis:

**Claim 1** Let $x \in \mathbb{R}^L$ and suppose there exists $k \in \mathbb{N}$ such that, for some $\alpha_1, \ldots, \alpha_k \in \mathbb{N}$ which are not all a multiple of 2, it holds $L = 3^{\alpha_1} p_2^{\alpha_2} \ldots p_k^{\alpha_k}$, with $p_2 < \ldots < p_k$ being prime numbers. Let also $\mathcal{G}_{a,b}^L$ denote the set of all spark deficient Gabor transforms generated according to Theorem 1, with respect to some lattice parameters $a$, $b$. Then, among all possible choices for $a$, $b$, the solution returned by (4) with $\Phi \in \mathcal{G}_{p_{k-1},p_k}^L$ is a global minimizer.

## 6 Conclusion and Future Directions

In the present paper, we took advantage of a window vector to generate a spark deficient Gabor frame and introduced a redundant analysis Gabor operator/DGT, namely star-DGT, associated with this SDGF. We then applied the star-DGT to analysis denoising, along with three other DGTs generated by state-of-the-art window vectors in the field of Gabor Analysis. First, we fixed the ambient dimension and the time-frequency parameters, and altered the standard deviation of the Gaussian noise and the amplitude of the coloured noises. Second, we examined how different pairs of lattice parameters, with fixed standard deviation and amplitude of the Gaussian and coloured noises respectively, affect the performance of analysis denoising. All experiments

confirm improved robustness: the increased amount of linear dependencies provided by this SDGF, yields for all speech signals a lower MSE for the proposed method. Future directions will include the combination of the present framework with deep learning methods [59], as well as the development of a mathematical framework, explaining why star-DGT benefits more when the lattice parameters are chosen as the two largest primes in the prime factorization of a signal's dimension (cf. Claim 1). Finally, it would be interesting to calculate – following research directions similar to [60] – non-trivial upper/lower bounds for the spark of our proposed SDGF.

## Acknowledgments

V. Kouni was financially supported for this work by the German Academic Exchange Service (DAAD) through the program *Research Grants - One-Year Grants for Doctoral Candidates, 2020-2021*. V. Kouni would also like to thank G. Paraskevopoulos for his valuable advice and insightful discussions around the framework presented in this paper.

## Conflict of Interest Statement

On behalf of all authors, the corresponding author states that there is no conflict of interest.

## References

[1] Chowdhury, T.H., Poudel, K.N., Hu, Y.: Time-frequency analysis, denoising, compression, segmentation, and classification of PCG signals. IEEE Access **8**, 160882–160890 (2020)

[2] Yasuda, M., Koizumi, Y., Saito, S., Uematsu, H., Imoto, K.: Sound event localization based on sound intensity vector refined by dnn-based denoising and source separation. In: ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 651–655 (2020). IEEE

[3] Grozdić, D.T., Jovičić, S.T., Subotić, M.: Whispered speech recognition using deep denoising autoencoder. Engineering Applications of Artificial Intelligence **59**, 15–22 (2017)

[4] Han, K., Wang, Y., Wang, D., Woods, W.S., Merks, I., Zhang, T.: Learning spectral mapping for speech dereverberation and denoising. IEEE/ACM Transactions on Audio, Speech, and Language Processing **23**(6), 982–992 (2015)

[5] Yu, C., Zezario, R.E., Wang, S.-S., Sherman, J., Hsieh, Y.-Y., Lu, X., Wang, H.-M., Tsao, Y.: Speech enhancement based on denoising autoencoder with multi-branched encoders. IEEE/ACM Transactions on Audio, Speech, and Language Processing **28**, 2756–2769 (2020)

[6] Zengyuan, L., Anming, D.: A speech denoising algorithm based on harmonic regeneration. In: IOP Conference Series: Earth and Environmental Science, vol. 332, p. 022042 (2019). IOP Publishing

[7] Grais, E.M., Plumbley, M.D.: Single channel audio source separation using convolutional denoising autoencoders. In: 2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP), pp. 1265–1269 (2017). IEEE

[8] Févotte, C., Torrésani, B., Daudet, L., Godsill, S.J.: Sparse linear regression with structured priors and application to denoising of musical audio. IEEE Transactions on Audio, Speech, and Language Processing **16**(1), 174–185 (2007)

[9] Attias, H., Platt, J.C., Acero, A., Deng, L.: Speech denoising and dereverberation using probabilistic models. In: Advances in Neural Information Processing Systems, pp. 758–764 (2001)

[10] Hasan, T., Hasan, M.K.: Suppression of residual noise from speech signals using empirical mode decomposition. IEEE Signal Processing Letters **16**(1), 2–5 (2008)

[11] Hussein, R., Shaban, K.B., El-Hag, A.H.: Denoising different types of acoustic partial discharge signals using power spectral subtraction. High voltage **3**(1), 44–50 (2018)

[12] Kamath, S., Loizou, P., *et al.*: A multi-band spectral subtraction method for enhancing speech corrupted by colored noise. In: ICASSP, vol. 4, pp. 44164–44164 (2002). Citeseer

[13] Yu, G., Mallat, S., Bacry, E.: Audio denoising by time-frequency block thresholding. IEEE Transactions on Signal processing **56**(5), 1830–1839 (2008)

[14] Siedenburg, K., Dörfler, M.: Audio denoising by generalized time-frequency thresholding. In: Audio Engineering Society Conference: 45th International Conference: Applications of Time-Frequency Processing in Audio (2012). Audio Engineering Society

[15] Rethage, D., Pons, J., Serra, X.: A wavenet for speech denoising. In: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 5069–5073 (2018). IEEE

[16] Xu, L., Choy, C.-S., Li, Y.-W.: Deep sparse rectifier neural networks for speech denoising. In: 2016 IEEE International Workshop on Acoustic Signal Enhancement (IWAENC), pp. 1–5 (2016)

[17] Masuyama, Y., Yatabe, K., Oikawa, Y.: Low-rankness of complex-valued spectrogram and its application to phase-aware audio processing. In: International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 855–859 (2019). IEEE

[18] Sprechmann, P., Bronstein, A., Bronstein, M., Sapiro, G.: Learnable low rank sparse models for speech denoising. In: International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 136–140 (2013). IEEE

[19] Plumbley, M.D., Blumensath, T., Daudet, L., Gribonval, R., Davies, M.E.: Sparse representations in audio and music: from coding to source separation. Proceedings of the IEEE **98**(6), 995–1005 (2009)

[20] Brajović, M., Stanković, I., Daković, M., Stanković, L.: Audio signal denoising based on laplacian filter and sparse signal reconstruction. In: 26th International Conference on Information Technology (IT), pp. 1–4 (2022). IEEE

[21] Liu, H., Liu, S., Li, Y., Li, D., Truong, T.-K.: Speech denoising based on group sparse representation in the case of gaussian noise. In: 23rd International Conference on Digital Signal Processing (DSP), pp. 1–5 (2018). IEEE

[22] Hadhami, I., Bouzid, A.: Speech denoising based on empirical mode decomposition and improved thresholding. In: International Conference on Nonlinear Speech Processing, pp. 200–207 (2013). Springer

[23] Abdulatif, S., Armanious, K., Guirguis, K., Sajeev, J.T., Yang, B.: Aegan: Time-frequency speech denoising via generative adversarial networks. In: 28th European Signal Processing Conference (EUSIPCO), pp. 451–455 (2021). IEEE

[24] Fletcher, A.K., Rangan, S., Goyal, V.K., Ramchandran, K.: Analysis of denoising by sparse approximation with random frame asymptotics. In: Proceedings. International Symposium on Information Theory, 2005. ISIT 2005., pp. 1706–1710 (2005). IEEE

[25] Coifman, R.R., Donoho, D.L.: Translation-invariant de-noising. In: Wavelets and Statistics, pp. 125–150. Springer, New York, NY (1995)

[26] Yatabe, K., Oikawa, Y.: Phase corrected total variation for audio signals. In: International Conference on Acoustics, Speech and Signal Processing

(ICASSP), pp. 656–660 (2018). IEEE

[27] Gaultier, C., Kitić, S., Bertin, N., Gribonval, R.: AUDASCITY: Audio denoising by adaptive social cosparsity. In: 25th European Signal Processing Conference (EUSIPCO), pp. 1265–1269 (2017). IEEE

[28] Genzel, M., Kutyniok, G., März, M.: $l_1$-analysis minimization and generalized (co-) sparsity: When does recovery succeed? Applied and Computational Harmonic Analysis **52**, 82–140 (2021)

[29] Selesnick, I.W., Figueiredo, M.A.: Signal restoration with overcomplete wavelet transforms: Comparison of analysis and synthesis priors. In: Wavelets XIII, vol. 7446, p. 74460 (2009). International Society for Optics and Photonics

[30] Kabanava, M., Rauhut, H.: Analysis $l_1$-recovery with frames and Gaussian measurements. Acta Applicandae Mathematicae **140**(1), 173–195 (2015)

[31] Elad, M.: Sparse and redundant representations: from theory to applications in signal and image processing (2010)

[32] Bhattacharya, G., Depalle, P.: Sparse denoising of audio by greedy time-frequency shrinkage. In: International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2898–2902 (2014). IEEE

[33] Lawrence, J., Pfander, G.E., Walnut, D.: Linear independence of Gabor systems in finite dimensional vector spaces. Journal of Fourier Analysis and Applications **11**(6), 715–726 (2005)

[34] Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D: nonlinear phenomena **60**(1-4), 259–268 (1992)

[35] Nam, S., Davies, M.E., Elad, M., Gribonval, R.: The cosparse analysis model and algorithms. Applied and Computational Harmonic Analysis **34**(1), 30–56 (2013)

[36] Blumensath, T., Davies, M.E.: Sampling theorems for signals from the union of finite-dimensional linear subspaces. IEEE Transactions on Information Theory **55**(4), 1872–1882 (2009)

[37] Fickus, M., Mixon, D.G., Tremain, J.C.: Steiner equiangular tight frames. Linear algebra and its applications **436**(5), 1014–1027 (2012)

[38] van Schijndel, N.H., Houtgast, T., Festen, J.M.: Intensity discrimination of gaussian-windowed tones: Indications for the shape of the auditory frequency-time window. The Journal of the Acoustical Society of America

**105**(6), 3425–3435 (1999)

[39] Guenther, F.H., Hickok, G.: Role of the auditory system in speech production. Handbook of clinical neurology **129**, 161–175 (2015)

[40] Qiu, A., Schreiner, C.E., Escabí, M.A.: Gabor analysis of auditory midbrain receptive fields: spectro-temporal and binaural composition. Journal of neurophysiology **90**(1), 456–476 (2003)

[41] Necciari, T., Holighaus, N., Balazs, P., Pruša, Z., Majdak, P., Derrien, O.: Audlet filter banks: A versatile analysis/synthesis framework using auditory frequency scales. Applied Sciences **8**(1), 96 (2018)

[42] Kouni, V., Rauhut, H.: Spark deficient Gabor frame provides a novel analysis operator for compressed sensing. In: Mantoro, T., Lee, M., Ayu, M.A., Wong, K.W., Hidayanto, A.N. (eds.) Neural Information Processing, pp. 700–708. Springer, Cham (2021)

[43] Zauner, G.: Quantum designs. PhD thesis, University of Vienna Vienna (1999)

[44] Zhivomirov, H.: A method for colored noise generation. Romanian journal of acoustics and vibration **15**(1), 14–19 (2018)

[45] Chardon, G., Necciari, T., Balazs, P.: Perceptual matching pursuit with Gabor dictionaries and time-frequency masking. In: International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 3102–3106 (2014). IEEE

[46] Becker, S.R., Candès, E.J., Grant, M.C.: Templates for convex cone problems with applications to sparse signal recovery. Mathematical programming computation **3**(3), 165 (2011)

[47] Søndergaard, P.L.: Efficient algorithms for the discrete Gabor transform with a long fir window. Journal of Fourier Analysis and Applications **18**(3), 456–470 (2012)

[48] Malikiosis, R.-D.: A note on Gabor frames in finite dimensions. Applied and Computational Harmonic Analysis **38**(2), 318–330 (2015)

[49] Scherzer, O.: Handbook of Mathematical Methods in Imaging, (2010). Springer Science & Business Media

[50] Malikiosis, R.-D.: Spark deficient Gabor frames. Pacific Journal of Mathematics **294**(1), 159–180 (2018)

[51] Dang, H.B., Blanchfield, K., Bengtsson, I., Appleby, D.M.: Linear dependencies in Weyl–Heisenberg orbits. Quantum Information Processing

**12**(11), 3449–3475 (2013)

[52] Søndergaard, P.L., Torrésani, B., Balazs, P.: The linear time frequency analysis toolbox. International Journal of Wavelets, Multiresolution and Information Processing **10**(04), 1250032 (2012)

[53] Panayotov, V., Chen, G., Povey, D., Khudanpur, S.: Librispeech: an ASR corpus based on public domain audio books. In: 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 5206–5210 (2015). IEEE

[54] Dahlke, S., Heuer, S., Holzmann, H., Tafo, P.: Statistically optimal estimation of signals in modulation spaces using Gabor frames. IEEE Transactions on Information Theory **68**(6), 4182–4200 (2022). IEEE

[55] Booth, T.E.: Power iteration method for the several largest eigenvalues and eigenfunctions. Nuclear science and engineering **154**(1), 48–62 (2006)

[56] Isar, D., Gajitzki, P.: Pink noise generation using wavelets. In: 2016 12th IEEE International Symposium on Electronics and Telecommunications (ISETC), pp. 261–264 (2016). IEEE

[57] Kailkhura, B., Thiagarajan, J.J., Bremer, P.-T., Varshney, P.K.: Stair blue noise sampling. ACM Transactions on Graphics (TOG) **35**(6), 1–10 (2016)

[58] Chergui, L., Bouguezel, S.: A new pre-whitening transform domain lms algorithm and its application to speech denoising. Signal processing **130**, 118–128 (2017)

[59] Luan, S., Chen, C., Zhang, B., Han, J., Liu, J.: Gabor convolutional networks. IEEE Transactions on Image Processing **27**(9), 4357–4366 (2018)

[60] Tillmann, A.M.: Computing the spark: mixed-integer programming for the (vector) matroid girth problem. Computational Optimization and Applications **74**(2), 387–441 (2019)