Simultaneous Optimized Orthogonal Matching Pursuit with Application to ECG Compression

Laura Rebollo-Neira
Department of Applied Mathematics and Data Science
Aston University
B3 7ET, Birmingham, UK

Abstract

A greedy pursuit strategy which finds a common basis for approximating a set of similar signals is proposed. The strategy extends the Optimized Orthogonal Matching Pursuit approach to selecting the subspace containing the approximation of all the signals in the set. The method, called Simultaneous Optimized Orthogonal Matching Pursuit, is stepwise optimal in the sense of minimizing at each iteration the mean error norm of the joint approximation. When applied to compression of electrocardiograms, significant gains over other transformation based compression techniques are demonstrated on the MIT-BIH Arrhythmia dataset.

Keywords: Simultaneous Optimized Orthogonal Matching Pursuit; Sparse Representation; ECG compression.

1 Introduction

Important signals in everyday life such as natural images, audio, and electrocardiogram records, are in general highly compressible. This implies that the original signal, available as a large set of numerical values, can be transformed into a set of much smaller cardinality or a set containing a large proportion of zero values. The transformation, which should not compromise the informational content of the data, is frequently called sparse representation. Traditional methods for sparse representation of signals are realized by applying an orthogonal transformation and disregarding the least relevant points in the transformed domain. Subsequently the signal is recovered by means of the inverse transformation. However, alternative transformations, which are not orthogonal but adapted to a signal at hand, have been shown to render high level of sparsity. Such transformations aim at representing a signal as a superposition of elements, which are called 'atoms' and are selected from a

large set called 'dictionary'. The superposition is said to be sparse if it involves a number of atoms much smaller than the number of numerical values representing the original signal.

Given a dictionary, the problem of finding the sparsest approximation of a signal, up to some acceptable error, is an NP-hard problem [1]. In practice it is addressed by tractable methodologies known as Pursuit Strategies. Such methodologies can be grouped for the most part in two broad categories. Namely, Basis Pursuit and Greedy Pursuit Strategies. The Basis Pursuit (BP) approach endeavors to obtain a tractable sparse solution by minimization of the 1-norm [2]. Greedy algorithms seek for a sparse solution by stepwise selection of dictionary's atoms. When dealing with real data the latter are in general more convenient. From the seminal Matching Pursuit (MP) [3] and Orthogonal Matching Pursuit (OMP) [4] methods, a number of Greedy Pursuit Strategies have been developed to improve the process of sparsely representing single signals [5–15]. Due to complexity issues and memory requirements, most of these techniques are to be applied by segmenting the signal and approximating each segment independently of the others. Nonetheless, when the segments bear similarity with each other, for some applications it is convenient to look for the dictionary's atoms suitable to represent all the segments simultaneously. The greedy Pursuit Strategy which has been dedicated to simultaneously approximate a set of signal is based on OMP [4] and has been termed Simultaneous Orthogonal Matching Pursuit (SOMP) [16]. Since in this work we extend the Optimized Orthogonal Matching Pursuit method [6] to simultaneously approximate a set of signals, we term the new approach Simultaneous Optimized Orthogonal Matching Pursuit (SOOMP).

The difference between SOMP and the SOOMP approach introduced in this work is equivalent to the difference between OMP and OOMP methods, both for approximating single signals. OOMP is stepwise optimal in the sense of minimizing at each iteration the norm of the residual error. Whilst OMP minimizes the norm of the error only with respect to the coefficients of the atomic superposition, OOMP minimizes the norm of the error with respect to those coefficients *and* the selection of a new atom. In the case of multiple signals SOOMP is designed to minimize the mean

value error norm. An additional advantage arises from the proposed implementation. Based on adaptive biorthogonalization, the SOOMP method produces at each iteration the common dual basis to the basis of selected atoms. This allows to calculate the coefficients of the representation of each signal in the set simply by computation of inner products. The practical relevance of the approach is illustrated by using it for compression of electrocardiogram (ECG) records.

An ECG signal represents a sequence of heartbeats which, if properly segmented and aligned, are suitable to be simultaneously approximated. This property is shown to benefit compression. Reliable comparison with other compression techniques is made possible by recuse to an adaptive quantization procedure that facilitates to reconstruct the whole ECG record at the required quality. The compression results are shown to significantly improve upon results produced by different transformation based approaches.

The paper is organized as follows: Sec. 2 introduces the problem and the mathematical notation. Sec. 3 establishes the proposed SOOMP approach for simultaneous approximation of a set of similar signals. Sec. 4 applies the method for compressing digital ECG records and produces reliable comparisons with previously reported results. The conclusions are presented in Sec. 5.

2 Mathematical introduction of the problem

In order to pose in mathematical terms the problem to be addressed we need to introduce the notation used throughout the paper as well as some preliminary background.

The sets of real, integer, and natural numbers are indicated by $\mathbb{R}, \mathbb{Z}$, and $\mathbb{N}$, respectively. Boldface letters are used to indicate Euclidean vectors or matrices whilst standard mathematical fonts indicate components, e.g., $\mathbf{f} \in \mathbb{R}^N$, $N \in \mathbb{N}$ is a vector of components $f(i)$, $i = 1, \ldots, N$ and $\mathbf{C} \in \mathbb{R}^{Q \times k}$ is a matrix of elements $C(i, j)$, $i = 1, \ldots, Q$, $j = 1, \ldots, k$ which when not leaving room for ambiguity will also be represented as $C(:, j), j = 1, \ldots, k$ A set of Q signals of equal length, say L, to be simultaneously approximated in a common subspace, is represented as a set of vectors $\{\mathbf{f}\{q\} \in$

$\mathbb{R}^L$, $q = 1, \ldots, Q\}$. The inner product is indicated as $\langle \cdot, \cdot \rangle$, e.g for $\mathbf{f}\{1\} \in \mathbb{R}^L$ and $\mathbf{f}\{2\} \in \mathbb{R}^L$

$$\langle \mathbf{f}\{1\}, \mathbf{f}\{2\} \rangle = \sum_{i=1}^{L} f\{1\}(i) f\{2\}(i).$$

The 2-norm induced by the inner product is denoted as $\| \cdot \|_2$, e.g. for $\mathbf{f}\{q\} \in \mathbb{R}^L$

$$\|\mathbf{f}\{q\}\|_2 = \sqrt{\langle \mathbf{f}\{q\}, \mathbf{f}\{q\} \rangle} = \sqrt{\sum_{i=1}^{L} \left(f\{q\}(i)\right)^2}.$$

A set of M vectors

$$\mathcal{D} = \left\{ \mathbf{d}_n \in \mathbb{R}^L \, ; \|\mathbf{d}_n\|_2 = 1 \right\}_{n=1}^{M},$$

such that $\text{span}\,(\mathcal{D}) = V_k$ and $\dim(V_k) \leq M$, is called a *dictionary* for V_k and its elements are called *atoms*.

In our context a signal $\mathbf{f}$ is considered to be an element of some inner product space V. Each given signal is assumed to be well approximated by an element, say $\mathbf{f}_k$, belonging to a finite dimensional subspace $V_k \subset V$. This assumption implies that, within a tolerance ρ much larger than the numerical errors in the calculations, $\mathbf{f}_k \in V_k$ is accepted to be a good approximation of $\mathbf{f} \in V$ if $\|\mathbf{f} - \mathbf{f}_k\|_2^2 < \rho$. Examples of signals fulfilling this definition are, amongst others, images, audio signals, and electrocardiograms. These are all signals with acceptable approximations which, without affecting their informational content, do not necessarily produce a highly accurate point-wise reproduction of the signals. These type of signals are suitable for lossy compression.

Since this work concerns approximation of similar signals we need to make an assumption on the signals that will be considered. We say that a finite set of Q signals $\{\mathbf{f}\{q\} \in V\}_{q=1}^{Q}$ are similar if they can be well approximated in a subspace V_k of dimension k, with k significantly smaller than the dimention of V. This is equivalent to assuming that there exists a common basis $\{\boldsymbol{\alpha}_n\}_{n=1}^{k}$ for V_k such that each signal $\mathbf{f}\{q\}$ is approximated as

$$\tilde{\mathbf{f}}\{q\} = \sum_{n=1}^{K} c_n^q \boldsymbol{\alpha}_n, \quad q = 1, \ldots, Q.$$

The quality of the approximated set will be assessed in mean value

$$\overline{\mathcal{E}} = \sum_{q=1}^{Q} p(q) \|\mathbf{f}\{q\} - \tilde{\mathbf{f}}\{q\}\|_2^2,$$

where $p(q) \geq 0$ with $\sum_{q=1}^{Q} p(q) = 1$.

3 Strategy for simultaneous approximation of a set of signals

Given a set of similar signals $\{\mathbf{f}\{q\} \in \mathbb{R}^L\}_{q=1}^{Q}$ and a dictionary, the aim is to simultaneously approximate all the signals in the set $\{\mathbf{f}\{q\} \in \mathbb{R}^L\}_{q=1}^{Q}$ within a common subspace $V_k = \text{span}\left(\{\mathbf{d}_{\ell_n}\}_{n=1}^{k}\right)$. In other words, each signal $\mathbf{f}\{q\} \in \mathbb{R}^L$ is to be approximated as a k-term atomic superposition

$$\mathbf{f}\{q\}^k = \sum_{n=1}^{k} c\{q\}(n)\mathbf{d}_{\ell_n}, \quad q = 1, \ldots, Q, \tag{1}$$

where the atoms $\mathbf{d}_{\ell_n}$, $n = 1, \ldots, k$ in (1) are selected from the given dictionary according to the criterion of optimality that will be established by Proposition 1 in the next subsection. Let us suppose for the moment that these atoms are known. Assigning a weight $p(q) \geq 0$ to the signal $\mathbf{f}\{q\}$, with $\sum_{q=1}^{Q} p(q) = 1$, the coefficients $\mathbf{c}\{q\} \in \mathbb{R}^k$ in (1) are required to minimize the mean of the square norm of the errors in the approximation of the set of signals, i.e.

$$\mathbf{c}\{q\}, \ldots, \mathbf{c}\{Q\} = \underset{\mathbf{c}'\{q\}, \ldots, \mathbf{c}'\{Q\}}{\arg\min} \sum_{q=1}^{Q} p(q) \|\mathbf{f}\{q\} - \sum_{n=1}^{k} c'\{q\}(n)\mathbf{d}_{\ell_n}\|_2^2. \tag{2}$$

Since $p(q) \geq 0$ the above minimization is equivalent to finding the components $c\{q\}(n)$, $n = 1, \ldots, k$ of each vector $\mathbf{c}\{q\}$ such that

$$c\{q\}(1), \ldots, c\{q\}(n) = \underset{c'\{q\}(1), \ldots, c'\{q\}(n)}{\arg\min} \|\mathbf{f}\{q\} - \sum_{n=1}^{k} c'\{q\}(n)\mathbf{d}_{\ell_n}\|_2^2 \quad q = 1 \ldots, Q. \tag{3}$$

Accordingly, the minimization with respect to the coefficients in (1) can be implemented by adaptive biorthogonalization [17], as proposed within the OOMP algorithm for a single signal [6],

$$c\{q\}(n) = \left\langle \boldsymbol{\beta}_n^k, \mathbf{f}\{q\} \right\rangle, \quad q = 1, \ldots, Q, \tag{4}$$

with vectors $\boldsymbol{\beta}_n^k$ calculated as will be described in the next section.

The selection of the atoms $\mathbf{d}_{\ell_n}$, $n = 1 \ldots, k$ in the decomposition (1) such that

$$\sum_{q=1}^{Q} p(q) \|\mathbf{f}\{q\} - \sum_{n=1}^{k} c\{q\}(n)\mathbf{d}_{\ell_n}\|_2^2 \text{ is minimized}$$

poses an intractable problem (for a dictionary of M atoms there are $\frac{M!}{(M-k)!k!}$ possibilities to be checked). We address the selection in a tractable manner by extending the OOMP strategy to simultaneously approximate a set of similar signals. The extended strategy is refereed to as SOOMP (Simultaneous OOMP).

3.1 SOOMP algorithm

The algorithm is initialized by setting: $\mathbf{r}\{q\}^0 = \mathbf{f}\{q\}$, $\mathbf{f}\{q\}^0 = 0$, $\Gamma = \emptyset$ and $k = 0$. The first atom is selected as the one corresponding to the index ℓ_1 such that

$$\ell_1 = \arg\max_{n=1,\ldots,M} \sum_{q=1}^{Q} p(q) \left|\left\langle \mathbf{d}_n, \mathbf{r}\{q\}^0 \right\rangle\right|^2. \tag{5}$$

This first atom is used to assign $\mathbf{w}_1 = \boldsymbol{\beta}_1 = \mathbf{d}_{\ell_1}$, calculate $\mathbf{r}\{q\}^1 = \mathbf{f}\{q\} - \mathbf{d}_{\ell_1} \langle \mathbf{d}_{\ell_1}, \mathbf{f}\{q\} \rangle$ and iterate as prescribed below.

1) Upgrade the set $\Gamma \leftarrow \Gamma \cup \ell_{k+1}$, increase $k \leftarrow k + 1$, and select the index of a new atom for the approximation as

$$\ell_{k+1} = \arg\max_{\substack{n=1,\ldots,M \\ n \notin \Gamma}} \sum_{q=1}^{Q} p(q) \frac{\left|\left\langle \mathbf{d}_n, \mathbf{r}\{q\}^k \right\rangle\right|^2}{1 - \sum_{i=1}^{k} |\langle \mathbf{d}_n, \widetilde{\mathbf{w}}_i \rangle|^2}, \quad \text{with} \quad \widetilde{\mathbf{w}}_i = \frac{\mathbf{w}_i}{\|\mathbf{w}_i\|}. \tag{6}$$

2) Compute the corresponding new vector $\mathbf{w}_{k+1}$ as

$$\mathbf{w}_{k+1} = \mathbf{d}_{\ell_{k+1}} - \sum_{i=1}^{k} \frac{\mathbf{w}_i}{\|\mathbf{w}_i\|^2} \left\langle \mathbf{w}_i, \mathbf{d}_{\ell_{k+1}} \right\rangle, \tag{7}$$

including for numerical accuracy the re-orthogonalization step:

$$\mathbf{w}_{k+1} \leftarrow \mathbf{w}_{k+1} - \sum_{i=1}^{k} \frac{\mathbf{w}_i}{\|\mathbf{w}_i\|^2} \left\langle \mathbf{w}_i, \mathbf{w}_{k+1} \right\rangle. \tag{8}$$

3) Upgrade vectors $\boldsymbol{\beta}_n^k$ as

$$\boldsymbol{\beta}_{k+1}^{k+1} = \frac{\mathbf{w}_{k+1}}{\|\mathbf{w}_{k+1}\|^2}, \quad \boldsymbol{\beta}_n^{k+1} = \boldsymbol{\beta}_n^k - \boldsymbol{\beta}_{k+1}^{k+1} \left\langle \mathbf{d}_{\ell_{k+1}}, \boldsymbol{\beta}_n^k \right\rangle, \quad n = 1, \ldots, k. \tag{9}$$

4) Update $\mathbf{r}\{q\}^k$ as

$$\mathbf{r}\{q\}^{k+1} \;=\; \mathbf{r}\{q\}^k - \langle \mathbf{w}_{k+1}, \mathbf{f}\{q\} \rangle \frac{\mathbf{w}_{k+1}}{\|\mathbf{w}_{k+1}\|^2}. \tag{10}$$

5) If for a given value ρ the condition $\sum_{q=1}^{Q} p(q) \|\mathbf{r}\{q\}^{k+1}\|_2^2 < \rho$ has been met stop the selection process. Otherwise repeat steps 1) - 5).

Once the iterations have finished calculate the coefficients for the decomposition (1) as

$$\mathbf{c}\{q\}_n = \left\langle \boldsymbol{\beta}_n^k, \mathbf{f}\{q\} \right\rangle, \quad n = 1, \ldots, k,\ q = 1, \ldots, Q.$$

For $q = 1, \ldots, Q$ calculate the final approximation of each signal $\mathbf{f}\{q\}$ as

$$\mathbf{f}\{q\}^k = \mathbf{f}\{q\} - \mathbf{r}\{q\}^k.$$

Remark 1. *The set of vectors $\boldsymbol{\beta}_n^k$, $n = 1, \ldots, k$, $q = 1, \ldots, Q$ as given in* (9) *fulfills that*

$$\mathbf{f}\{q\}^k = \sum_{n=1}^{k} \left\langle \beta_n^k, \mathbf{f}\{q\} \right\rangle \mathbf{d}_{\ell_n} = \hat{P}_{V_k} \mathbf{f}\{q\}, \quad q = 1, \ldots, Q,$$

where $\hat{P}_{V_k}\mathbf{f}\{q\}$ is the orthogonal projector of $\mathbf{f}\{q\}$ onto $V_k = \operatorname{span}\{\mathbf{d}_{\ell_n}\}_{n=1}^k$. Please find the proof in [6], or as a particular case of the more general proof in [17].

Proposition 1. *The recursive selection of the indices $\ell_1, \ldots, \ell_k$, as proposed in* (6)*, is stepwise optimal. It minimizes, at each iteration, the mean of the square distance between the set of signals $\mathbf{f}\{q\}$, $q = 1, \ldots, Q$ and their corresponding approximations $\mathbf{f}\{q\}^k$, $q = 1, \ldots, Q$.*

Proof. For $k = 0$ it is clear that ℓ_1 selected as in (5) minimizes the mean of the square distance $\overline{\mathcal{E}^1}$ as given by

$$\overline{\mathcal{E}^1} = \sum_{q=1}^{Q} p(q) \|\mathbf{f}\{q\} - \mathbf{f}\{q\}^1\|_2^2 = \sum_{q=1}^{Q} p(q) (\|\mathbf{f}\{q\}\|_2^2 - |\langle \mathbf{d}_{\ell_1}, \mathbf{f}\{q\} \rangle|^2).$$

Let us assume that the indices $\ell_1, \ldots, \ell_k$ selected as proposed in (6) minimize, in the specified stepwise sense, the mean square distance

$$\overline{\mathcal{E}^k} = \sum_{q=1}^{Q} p(q) \|\mathbf{f}\{q\} - \mathbf{f}\{q\}^k\|_2^2.$$

We shall prove by induction that if the atoms $\mathbf{d}_{\ell_1}, \ldots, \mathbf{d}_{\ell_k}$ are fixed, at iteration $k+1$ the atom $\mathbf{d}_{\ell_{k+1}}$ selected as in (6) minimizes $\overline{\mathcal{E}^{k+1}}$. The proof stems from the fact that at iteration k the approximation $\mathbf{f}\{q\}^k$ of each signal $\mathbf{f}\{q\}$ is the orthogonal projection of $\mathbf{f}\{q\}$ onto the subspace $V_k = \text{span}\{\mathbf{d}_{\ell_n}\}_{n=1}^k$ (c.f. Remark 1).

Consider that V_k is augmented by one element, say $\mathbf{d}_{\ell_{k+1}} \notin V_k$, so that $V_{k+1} = V_k \oplus \mathbf{d}_{\ell_{k+1}}$, where $\oplus$ indicates direct sum. The orthogonal projection of each signal $\mathbf{f}\{q\}$, $q = 1, \ldots, Q$ onto V_{k+1} can be expressed as

$$\mathbf{f}\{q\}^{k+1} = \hat{P}_{V_{k+1}} \mathbf{f}\{q\} = \hat{P}_{V_k} \mathbf{f}\{q\} + \frac{\mathbf{w}_{k+1}}{\|\mathbf{w}_{k+1}\|_2^2} \langle \mathbf{w}_{k+1}, \mathbf{f}\{q\} \rangle \quad \text{with} \quad \mathbf{w}_{k+1} = \mathbf{d}_{\ell_{k+1}} - \hat{P}_{V_k} \mathbf{d}_{\ell_{k+1}}.$$

Thus

$$\begin{aligned} \|\mathbf{f}\{q\} - \mathbf{f}\{q\}^{k+1}\|_2^2 &= \|\mathbf{f}\{q\} - \hat{P}_{V_{k+1}} \mathbf{f}\{q\}\|_2^2 \\ &= \|\mathbf{f}\{q\} - \hat{P}_{V_k} \mathbf{f}\{q\} - \frac{\mathbf{w}_{k+1}}{\|\mathbf{w}_{k+1}\|_2^2} \langle \mathbf{w}_{k+1}, \mathbf{f}\{q\} \rangle \|_2^2 \\ &= \|\mathbf{f}\{q\} - \hat{P}_{V_k} \mathbf{f}\{q\}\|_2^2 - \frac{|\langle \mathbf{w}_{k+1}, \mathbf{f}\{q\} \rangle|_2^2}{\|\mathbf{w}_{k+1}\|_2^2}. \end{aligned}$$

Since $\|\mathbf{f}\{q\} - \hat{P}_{V_k} \mathbf{f}\{q\}\|_2^2$ is optimized and fixed at iteration k, it is true that at iteration $k+1$ the index of the atom which minimizes $\overline{\mathcal{E}^{k+1}}$ fulfils

$$\begin{aligned} \ell_{k+1} &= \underset{\substack{n=1,\ldots,M \\ n \notin \Gamma}}{\arg\max} \sum_{q=1}^{Q} p(q) \frac{|\langle \mathbf{w}_{k+1}, \mathbf{f}\{q\} \rangle|_2^2}{\|\mathbf{w}_{k+1}\|_2^2} \\ &= \underset{\substack{n=1,\ldots,M \\ n \notin \Gamma}}{\arg\max} \sum_{q=1}^{Q} p(q) \frac{\left| \left\langle \mathbf{d}_{\ell_n} - \hat{P}_{V_k} \mathbf{d}_{\ell_n}, \mathbf{f}\{q\} \right\rangle \right|^2}{\|\mathbf{d}_{\ell_n} - \hat{P}_{V_k} \mathbf{d}_{\ell_n}\|_2^2}. \end{aligned} \tag{11}$$

The proof is concluded using the self-adjoint properties of $\hat{P}_{V_k}$ to write:

$$\begin{aligned}\left|\left\langle \mathbf{d}_{\ell_n} - \hat{P}_{V_k}\mathbf{d}_{\ell_n}, \mathbf{f}\{q\}\right\rangle\right| &= \left|\langle \mathbf{d}_{\ell_n}, \mathbf{f}\{q\}\rangle - \left\langle \hat{P}_{V_k}\mathbf{d}_{\ell_n}, \mathbf{f}\{q\}\right\rangle\right| \\ &= \left|\left\langle \mathbf{d}_{\ell_n}, \mathbf{f}\{q\} - \hat{P}_{V_k}\mathbf{f}\{q\}\right\rangle\right| \\ &= \left|\left\langle \mathbf{d}_{\ell_n}, \mathbf{r}\{q\}^k\right\rangle\right|.\end{aligned}$$

Moreover, since all atoms are normalized and the set $\{\widetilde{\mathbf{w}}_i\}_{i=1}^k$ is an orthonormal basis for V_k we have

$$\hat{P}_{V_k}\mathbf{d}_{\ell_n} = \sum_{i=1}^{k} \widetilde{\mathbf{w}}_i \langle \widetilde{\mathbf{w}}_i, \mathbf{d}_{\ell_n}\rangle,$$

so that

$$\|\mathbf{d}_{\ell_n} - \hat{P}_{V_k}\mathbf{d}_{\ell_n}\|_2^2 = 1 - \|\hat{P}_{V_k}\mathbf{d}_{\ell_n}\|_2^2 = 1 - \sum_{i=1}^{k} |\langle \widetilde{\mathbf{w}}_i, \mathbf{d}_{\ell_n}\rangle|^2,$$

which shows the equivalence between (11) and (6). □

4 Application to compression of ECG records

A digital ECG signal represents a sequence of heartbeats. In a typical record each heartbeat is characterized by a combination of three graphical deflections, known as QRS complex, and two lateral and less visually noticeable P and T waves. A short segment of a typical ECG record is illustrated in Fig.1.

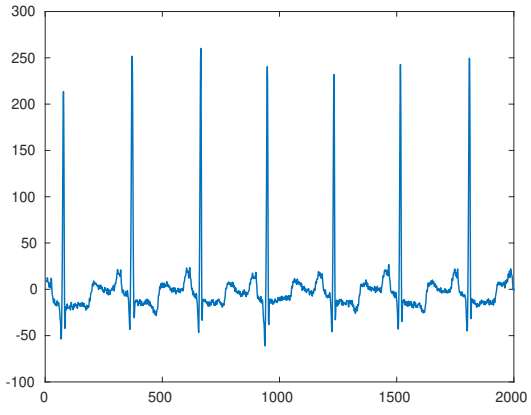

Figure 1: A short segment of an ECG record

In order to simultaneously approximate all the beats in a record we need to segment and align the beats to meet the requirement of being similar. The procedure is discussed in the next subsection.

4.1 Segmentation and alignment of heartbeats

The QRS complex is segmented once the central R peak is detected. This can be effectively done by the Pan Tompkins method [18]. In our numerical examples we use the off-the-shelf MATLAB implementation of this algorithm [19]. Since the distance between peaks in a record is not uniform, the length of the segmented beats should be passed to the decoder. The segmented peaks are placed in arrays $\mathbf{f}\{q\}$, $q = 1, \ldots, Q$ of equal length L by padding with zeros. Fig.4 illustrates the resulting configuration with 80 heartbeats.

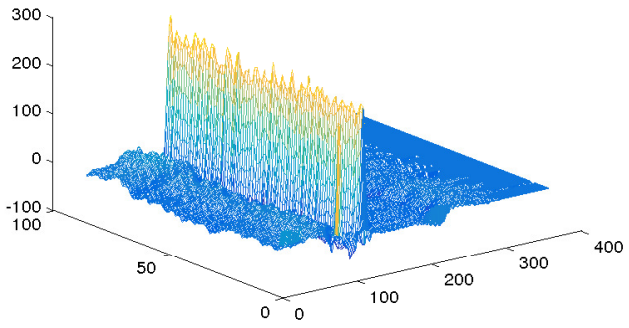

Figure 2: Configuration resulting by segmentation and alignment of 80 heartbeats for illustration purposes.

The segmented and aligned heartbeats are simultaneously approximated using the SOOMP approach by assigning the same weight to each heartbeat, i.e. setting $p(q) = \frac{1}{Q}$, $q = 1, \ldots, Q$. The adopted dictionary is the Cohen-Daubechies-Feauveau **CDF97** dictionary, of redundancy approximately two, introduced in [20, 21]. Given a partition x_i, $i = 1, \ldots, N$ of the interval $[c, d]$ the dictionary is constructed as follows [20, 21].

$$\mathcal{D} = \mathcal{V}_0 \cup \mathcal{W}_0 \cup \mathcal{W}_1 \cup \mathcal{W}_2 \cup \mathcal{W}_3 \cup \mathcal{W}_4, \tag{12}$$

with

$$\mathcal{V}_0 = \{\phi(x_i - \frac{k}{2})|_{[c,d]},\ k \in \mathbb{Z},\ i = 1, \ldots, N\}, \tag{13}$$

and

$$\mathcal{W}_j = \{2^{j/2}\psi(2^j x_i - \frac{k}{2})|_{[c,d]},\ k \in \mathbb{Z},\ i = 1, \ldots, N\}, \tag{14}$$

where $\psi(2^j x_i - \frac{k}{2})|_{[c,d]}$ indicates the restriction of the function $\psi(2^j x_i - \frac{k}{2})$ to the interval $[c, d]$. The prototype functions $\phi(x)$ and $\psi(x)$ are plotted in the left and right graphs of Fig.3 respectively. The MATLAB codes for producing numerically both functions and building the dictionary (12) are described in [21]. The codes have been made available in [23] together with of the complete MATLAB software for reproducing the numerical examples in this work.

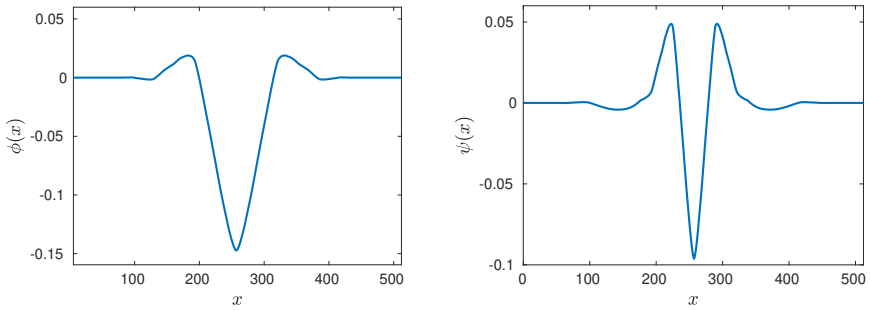

Figure 3: Cohen-Daubechies-Feauveau scaling and wavelet functions [22].

The left graph of Fig. 4 is the two dimensional image of the segmented and aligned heartbeats corresponding to the record 100 in the MIT-BIH Arrhythmia database [24].

The coefficients in the decompositions (1) are placed in a two dimensional array $\mathbf{C} \in \mathbb{R}^{Q \times k}$.

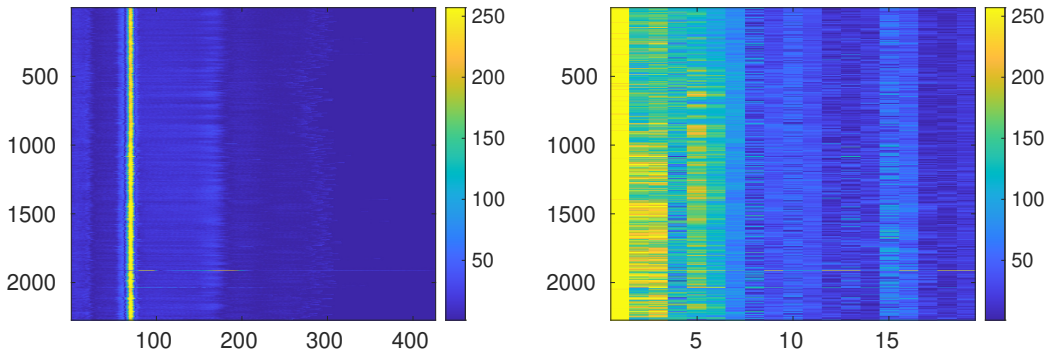

Figure 4: The graph on the left is the image of the magnitude of the aligned heartbeats in record 100. The graph on the right is the image of the magnitude of the coefficients in the simultaneous approximation of the heartbeats.

The right graph in Fig.4 shows the magnitude of the $\mathbf{C}$ entries corresponding to the approximation of the heartbeats represented in the left graph of the same figure. We notice that the size of the array in the left graph (2273×426) is reduced to a smaller array of size 2273×19 containing the coefficients in (1) with respect to the common basis $\mathbf{d}_{\ell_1}, \ldots \mathbf{d}_{\ell_k}$. As it is clear from the location of the brightest pixels in the right image of Fig.4 the coefficients of largest magnitude are all concentrated in the first vertical lines. This implies that to favor compression it is convenient to apply an orthogonal transformation to map the coefficients in the vertical direction to smaller values, which eventually will be quantized to zero. Thus, by applying the discrete cosine transform on each columns of $\mathbf{C}$, we create the transformed array $\mathbf{B} \in \mathbb{R}^{Q \times k}$ with the following entries

$$\mathbf{B}(:, n) = \widehat{\mathrm{dct}}\mathbf{C}(:, n), \quad n = 1, \ldots, k, \tag{15}$$

where $\widehat{\mathrm{dct}}\mathbf{C}(:, n)$ indicates the one dimensional discrete cosine transform operating on the n-th column of array $\mathbf{C}$. The magnitude of the transformed points is shown Fig. 5.

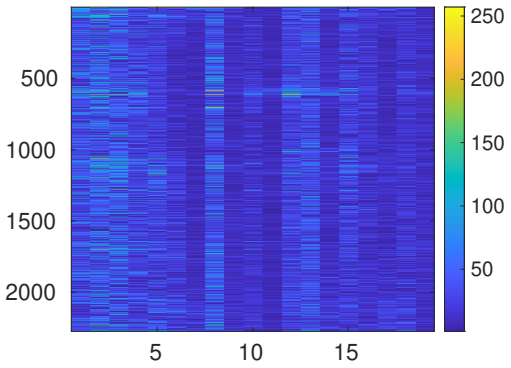

Figure 5: Magnitude of the array $\mathbf{B}$ (c.f. (15)).

4.2 Encoding

At the encoding step the $Q \times k$ array $\mathbf{B}$ is expressed as a vector $\mathbf{b} = (b(1), \ldots, b(K))$ of $K = Q \cdot k$ components, adopting the column-major order. The encoding of this vector follows the procedure outline in [25]. The components of $\mathbf{b}$ are converted to integer numbers by a mid-tread uniform

quantizer as follows:

$$b^{\Delta}(i) = \left\lfloor \frac{b(i)}{\Delta} + \frac{1}{2} \right\rfloor, \quad i = 1, \ldots, K, \tag{16}$$

where $\lfloor x \rfloor$ indicates the largest integer smaller or equal to x and Δ is the quantization parameter. For comparison with results in other publications in the numerical examples the quantization parameter Δ is set to produce the required quality of the reconstructed signal.

The absolute value of the elements (16) are placed in a smaller vector, say $\mathbf{b}' = (b'(1), \ldots, b'(K'))$, after the elimination of zeros. The signs are encoded separately in a vector $\mathbf{s} = (s(1), \ldots, s(K'))$ using a binary alphabet: 1 for + and 0 for −.

Assuming that the nonzero values in (16) occur at the positions $\jmath_i, \ldots, \jmath_{K'}$, these indices are re-ordered in ascending order $\jmath_i \to \tilde{\jmath}_i$, $i = 1, \ldots, K'$, i.e. $\tilde{\jmath}_i < \tilde{\jmath}_{i+1}$, $i = 1, \ldots, K'$. This induces new order in the coefficients, $\mathbf{b}' \to \tilde{\mathbf{b}}'$ and in the corresponding signs $\mathbf{s} \to \tilde{\mathbf{s}}$. Defining $\delta(i) = \tilde{\jmath}_i - \tilde{\jmath}_{i-1}$, $i = 2, \ldots, K'$ the array $\boldsymbol{\delta} = (\tilde{\jmath}_1, \delta(2), \ldots, \delta(K'))$ stores the indices $\tilde{\jmath}_1, \ldots, \tilde{\jmath}_{K'}$ with unique recovery.

Finally the vectors $\tilde{\mathbf{b}}', \tilde{\mathbf{s}}, \boldsymbol{\delta}$, as well as the length of the heartbeats $\mathbf{h}$, are compressed using adaptive Huffman coding implemented by the off-the-shelf MATLAB function Huff06 [26]. The additional numbers which have to be passed to the decoder are:

(i) The indices ℓ_i, $i = 1, \ldots, k$ of the selected dictionary's atoms forming the common basis.

(ii) The quantization parameter Δ.

(iii) The mean value of the 1D ECG record (if not previouly substracted).

(iv) The number of rows and columns of $\mathbf{C}$, i.e. Q and k.

4.3 1D ECG signal recovery

At the decoding stage, after reverting Huffman coding, the locations $\tilde{\jmath}_1, \ldots, \tilde{\jmath}_K$ of the nonzero entries in the transformed array after quantization are readily obtained. This allows the recovery of the array $\mathbf{B}^r$ as follows.

(i) Set $b^r(i) = 0$, $i = 1, \ldots, K$ and $b^r(\tilde{j}_i) = (2\tilde{s}(i) - 1)\tilde{b}'(i)\Delta$, $i = 1, \ldots, K$.

(ii) Reshape the vector $\mathbf{b}^r$ to produce a 2D array $\mathbf{B}^r$ of size $Q \times k$. The array $\mathbf{C}^r$ is recovered from the $\mathbf{B}^r$ one by inverting the $\widehat{\text{dct}}$ transformation (c.f. (15)).

(iii) Each row of the recovered array $\mathbf{C}^r$ gives the coefficients in the decomposition (1) of the approximated heartbeats, i.e. $\mathbf{f}^{\text{r}}\{q\} = \sum_{i=1}^{k} \mathbf{C}^r(q, i)\mathbf{d}_{\ell_i}, \quad q = 1, \ldots, Q$.

(iv) Finally the reconstructed beats $\mathbf{f}^{\text{r}}\{q\}$ are assembled in a 1D record using the distance between heartbeats that was stored in the vector $\mathbf{h}$.

The achieved compression ratio CR, which is defined as

$$\text{CR} = \frac{\text{Size of the uncompressed file}}{\text{Size of the compressed file}}, \tag{17}$$

depends on the required quality of the recovered signal. In the numerical examples the quality of the recovered records is assessed by the PRDN metric defined as follows

$$\text{PRDN} = \frac{\|\mathbf{f} - \mathbf{f}^{\text{r}}\|}{\|\mathbf{f} - \bar{\mathbf{f}}\|} \times 100\% \tag{18}$$

where $\mathbf{f}$ is the whole ECG record, $\mathbf{f}^{\text{r}}$ is the reconstructed record from the compressed file and $\bar{\mathbf{f}}$ is the mean of $\mathbf{f}$. It is pertinent to stress the importance of adopting the normalized metric (18) for comparison of reconstruction quality. The subtraction of $\bar{\mathbf{f}}$ avoids dependence on the signal baseline.

5 Numerical Tests

For the numerical test we use the MIT-BIH Arrhythmia database [24]. Each of the records is of 30 min length, consisting of $N = 650000$ 11-bit samples at a frequency of 360 Hz.

For comparison purposes we compress the subset of records reported in [27], [28], and [29] and reproduce the values of PRDN in those publications. This is achieved as follows: The SOOMP method is applied to approximate the set of heartbeats in each record up to 80% the target PRDN.

The quantization parameter Δ is automatically fixed, by a bisection algorithm, in order to reproduce the target PRDN for the whole record within two decimal places.

The first, second and third columns of Table 1 reproduce the results published in [27]. The comparison is relevant because the approach [27] is also based on approximation of heartbeats using a dictionary. The techniques are very different though. Whilst our dictionary does not have to be stored because it is numerically generated, the dictionary in [27] is part of the ECG record to be compressed. Moreover, the method for finding the sparse representation is different and so is the procedure to store the parameters that should be passed to the decoder.

Our compression results are shown in the forth column of Table 1. These results demonstrate a significant gain in CR for the same recovery quality. For further comparison we apply the fast compression algorithm [25], which does not require peak segmentation or Huffman coding. This method has been already shown to improve the average CR for the 48 records in the MIT-BIH Arrhythmia dataset with respect to the results in [30], [31], and [32], for a broad rage of average qualities. For comparison with [27] in Table 1 the compression is realized to reproduce the PRDN listed in the second column for each record.

Table 1: Comparison with the results in [27]. The first collumn lists the records considered in [27]. The second column displays the values of PRDN and the third collum their CRs. Our CRs for the same PRDN are shown in the forth column. The fifth column shows the CRs obtained with the fast approach [25].

Record	PRDN	CR [27]	CR prop.	CR [25]
100	18.03	78.20	**143.99**	36.51
100	17.22	75.12	**139.47**	35.25
101	14.66	80.24	**102.58**	31.26
101	12.91	76.46	**82.31**	30.31
102	18.54	**58.54**	**58.49**	33.89
102	18.16	48.47	**58.13**	33.48
103	12.57	46.32	**90.91**	30.84
103	11.57	44.33	**86.27**	29.61
109	13.70	24.86	**145.80**	51.23
109	9.97	23.53	**97.73**	36.91
111	26.20	31.05	**121.09**	38.29
111	19.51	29.44	**60.49**	32.20
112	16.58	34.06	**91.48**	35.05
112	15.99	35.49	**85.59**	34.32
113	14.08	37.42	**90.76**	32.49
113	9.82	32.55	**55.30**	27.68
115	9.76	38.26	**62.31**	24.52
115	9.18	36.57	**57.32**	23.74
117	14.42	38.94	**120.89**	36.94
117	13.38	37.13	**105.97**	35.74
119	32.19	16.26	**153.33**	90.40
119	16.36	15.24	**78.81**	48.08
121	17.36	26.67	**111.74**	46.45
121	15.63	25.29	**100.72**	41.11
Average	17.33	41.9	**107.78**	40.65
Average	14.14	39.97	**84.00**	34.04

Table 2: Same description as in Table 1 but the comparison is with the results of Table I in [28].

Record	PRDN	CR [28]	CR prop.	CR [25]
100	11.46	39.81	**64.47**	23.03
101	14.13	42.04	**95.53**	30.91
102	19.94	41.09	**63.69**	35.09
103	6.72	**41.24**	39.17	21.05
107	13.27	41.84	**71.40**	38.15
109	7.31	38.25	**59.34**	28.46
111	13.94	**41.73**	**41.59**	25.14
115	8.04	42.71	**47.79**	22.04
117	10.00	46.75	**51.56**	26.17
118	15.33	39.60	**66.44**	28.26
119	9.67	41.97	**43.98**	30.71
213	13.63	32.58	**62.24**	25.72
222	22.44	**40.69**	31.95	24.77
232	20.76	42.28	**59.66**	35.52
Average	13.33	40.90	**57.09**	28.22

The first, second and third columns of Table 2 reproduce the results published in [28], which are achieved with an approach based on the Singular Value Decomposition (SVD).

Our compression ratios (CRs) are shown in the the forth column of Table 2. The fifth column shows the CRs produced by the fast compression algorithm [25].

Table 3: Same description as in Table 1 but the comparison is with the results of Table I in [29].

Record	PRDN	CR [29]	CR prop.	CR [25]
100	11.55	50.70	**65.47**	23.14
101	11.29	58.54	**64.13**	27.86
103	9.16	54.16	**61.37**	25.70
107	14.53	53.12	**78.31**	40.34
109	11.83	46.43	**120.08**	44.67
111	16.40	**53.39**	50.11	28.34
115	8.94	56.77	**58.76**	23.49
117	12.43	66.15	**90.39**	34.62
119	10.28	**56.03**	46.99	32.17
214	17.03	50.84	**87.66**	54.04
223	17.49	45.38	**86.06**	40.80
Avegave	12.81	53.77	**73.57**	34.11

The first, second and third columns of Table 3 reproduce the results published in [29], which

are also obtained with a Singular Value Decomposition based approach. Our CRs are shown in the forth column of this table. The fifth column shows the CRs produced by the fast compression algorithm [25].

Note: The MATLAB software for reproducing the tables is available on

`http://www.nonlinear-approx.info/examples/node017.html`

6 Conclusions

The Optimized Orthogonal Matching Pursuit approach has been extended with the purpose of selecting a common basis for the simultaneous approximation of a set of similar signals. The extended approach, termed Simultaneous Optimized Orthogonal Matching Pursuit, minimizes at each iteration the mean error norm of the joint approximation. The applicability of the method was illustrated by the simultaneous approximation of heartbeats in ECG records taken from the MIT-BIH Arrhythmia database. The particular records were selected for comparison purposes as in [27], [28], and [29]. It was demonstrated that simultaneous approximations of heartbeats can be used for compressing a whole record. The adopted compression strategy was shown to improve upon compression results achieved by other methods for the same reconstruction quality. The comparison was made possible by means of an iterative quantization procedure which delivers the required quality. We feel confident that the proposed technique could be of assistance to other signal processing applications which benefit from the use of a common basis for the approximation of a set of similar signals.

References

[1] B. K. Natarajan, "Sparse Approximate Solutions to Linear Systems", *SIAM Journal on Computing*, **24**, 227–234 (1995).

[2] S. S. Chen, D. L. Donoho, and M. A Saunders, "Atomic Decomposition by Basis Pursuit", *SIAM Journal on Scientific Computing*, **20**, 33–61 (1998).

[3] S. G. Mallat and Z. Zhang, "Matching Pursuits with Time-Frequency Dictionaries", *IEEE Trans. Signal Process.,* **41**, 3397–3415 (1993).

[4] Y.C. Pati, R. Rezaiifar, and P.S. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," *Proc. of the 27th ACSSC*,**1**, 40–44 (1993).

[5] R. Gribonval, "Fast Matching Pursuit with a multiscale dictionary of Gaussian Chirps", *IEEE Trans. Signal Process.*, **49**, 994–1001 (2001).

[6] L. Rebollo-Neira and D. Lowe, "Optimized orthogonal matching pursuit approach", *IEEE Signal Process. Letters*, **9**, 137–140 (2002).

[7] M. Andrle, L. Rebollo-Neira, E. Sagianos, "Backward-optimized orthogonal matching pursuit approach", *IEEE Signal Proc. Let.*,**11**, 705–708 (2004).

[8] M. Andrle and L. Rebollo-Neira, "A swapping-based refinement of orthogonal matching pursuit strategies", *Signal Processing*, **86**, 480–495 (2006).

[9] D. L. Donoho , Y. Tsaig , I. Drori , and J. Starck, "Stagewise Orthogonal Matching Pursuit", *IEEE Transactions on Information Theory*, **58**, 1094–1121 (2006).

[10] D. Needell and J.A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples", *Applied and Computational Harmonic Analysis*, **26**, 301–321 (2009).

[11] D. Needell and R. Vershynin, "Signal Recovery From Incomplete and Inaccurate Measurements via Regularized Orthogonal Matching Pursuit", *IEEE Journal of Selected Topics in Signal Processing,* **4**, 310–316 (2010).

[12] Y. Eldar, P. Kuppinger, and H. Biölcskei, "Block-Sparse Signals: Uncertainty Relations and Efficient Recovery", *IEEE Trans. Signal Process.*, **58**, 3042–3054 (2010).

[13] L. Rebollo-Neira and J.Bowley, Sparse representation of astronomical images, *Journal of The Optical Society of America A,* **30**, 758–768 (2013).

[14] L. Rebollo-Neira, R. Matiol, and S. Bibi, "Hierarchized block wise image approximation by greedy pursuit strategies," *IEEE Signal Process. Letters*, **20**, 1175–1178 (2013).

[15] L. Rebollo-Neira, "Cooperative Greedy Pursuit Strategies for Sparse Signal Representation by Partitioning", *Signal Processing*, **125**, 365–375 (2016).

[16] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit", *Signal Processing*, **86**, 572–588 (2006).

[17] L. Rebollo-Neira, "Constructive updating/downdating of oblique projectors: a generalisation of the Gram-Schmidt process", *J. Phys. A: Math. Theor.* **40**, 6381–6394 (2007).

[18] J. Pan and W.J. Tompkins,"A Real-Time QRS Detection Algorithm", *IEEE Trans. on Biomedical Engineering*, **BME-32**, 230–236 (1985).

[19] H. Sedghamiz, Complete Pan Tompkins Implementation ECG QRS detector (https://www.mathworks.com/matlabcentral/fileexchange/45840-complete-pan-tompkins-implementation-ecg-qrs-detector), MATLAB Central File Exchange. Retrieved June 4, 2024.

[20] L. Rebollo-Neira and D. Černá, "Wavelet based dictionaries for dimensionality reduction of ECG signals", *Biomedical Signal Processing and Control*, **54,** 101593 (2019).

[21] D. Černá and L. Rebollo-Neira, "Construction of wavelet dictionaries for ECG modelling", *MethodsX,* **8**, 101314 (2021).

[22] A. Cohen, I. Daubechies, and J.C. Feauveau, "Biorthogonal bases of compactly supported wavelets", *Comm. Pure and Appl. Math.*, **45**, 485–560 (1992).

[23] `http://www.nonlinear-approx.info/examples/node017.html`

[24] https://physionet.org/physiobank/database/mitdb/

[25] L. Rebollo-Neira, "Effective high compression of ECG signals at low level distortion", *Scientific Reports,* **9**, No 4564 (2019).

[26] Karl Skretting. Huffman Coding and Arithmetic Coding (https://www.mathworks.com/matlabcentral/fileexchange/2818-huffman-coding-and-arithmetic-coding), MATLAB Central File Exchange. Retrieved June 4, 2024.

[27] A. Adamo, G. Grossi, R. Lanzarotti, J. Lin, "ECG compression retaining the best natural basis k-coefficients via sparse decomposition", *Biomedical Signal Processing and Control,* **15**, 11–17 (2015).

[28] T. Yu Liu, K. Jen Lin, and H. Chun Wu, "ECG Data Encryption Then Compression Using Singular Value Decomposition", *IEEE Journal of Biomedical and Health Informatics*, **22**, 707 – 713 (2018).

[29] L. Zheng, Z. Wang, J. Liang, S. Luo, and S. Tian, "Effective compression and classification of ECG arrhythmia by singular value decomposition",*Biomedical Engineering Advances*, **2**, 100013 (2021).

[30] S.J. Lee, J. Kim, and M. Lee,"A Real-Time ECG Data Compression and Transmission Algorithm for an e-Health Device", *IEEE Trans. on Biomedical Engineering*, **58,** 2448 – 2455 (2011).

[31] J.L. Ma, T.T. Zhang, and M. C. Dong, "A Novel ECG Data Compression Method Using Adaptive Fourier Decomposition With Security Guarantee in e-Health Applications", *IEEE Journal of Biomedical and Health Informatics,* **19**, 986 – 994 (2015).

[32] C. Tan, L. Zhang and H. Wu,"A Novel Blaschke Unwinding Adaptive Fourier Decomposition based Signal Compression Algorithm with Application on ECG Signals", *IEEE Journal of Biomedical and Health Informatics* **23**, pp 672 – 682 (2018).