
Personalized Convolutional Dictionary Learning of Physiological Time Series

Axel Roques

Centre Borelli,

Laboratoire GBCM, Thales AVS

Samuel Gruffaz

Centre Borelli,

ENS Paris-Saclay

Kyurae Kim

University of Pennsylvania

Alain O. Durmus

CMAP, CNRS, Ecole polytechnique

Laurent Oudre

Centre Borelli, ENS Paris-Saclay

Abstract

Human physiological signals tend to exhibit both global and local structures: the former are shared across a population, while the latter reflect inter-individual variability. For instance, kinetic measurements of the gait cycle during locomotion present common characteristics, although idiosyncrasies may be observed due to biomechanical disposition or pathology. To better represent datasets with local-global structure, this work extends Convolutional Dictionary Learning (CDL), a popular method for learning interpretable representations, or *ictionaries*, of time-series data. In particular, we propose Personalized CDL (PerCDL), in which a *local* dictionary models local information as a personalized spatiotemporal transformation of a *global* dictionary. The transformation is learnable and can combine operations such as time warping and rotation. Formal computational and statistical guarantees for PerCDL are provided and its effectiveness on synthetic and real human locomotion data is demonstrated.

1 INTRODUCTION

Dictionary learning (DL; Kreutz-Delgado et al., 2003; Tošić and Frossard, 2011) is an unsupervised representation learning method for decomposing data into two

components: a *dictionary*, *i.e.*, a collection of reference signals, each called an *atom*, and a set of mixing coefficients. Among the numerous variants of DL, Convolutional Dictionary Learning (CDL; Garcia-Cardona and Wohlberg, 2018) assumes that the dataset can be synthesized with atoms of relatively smaller length repeated at different positions. CDL is a well-established method to represent various time-series modalities (*e.g.*, audio signals; Grosse et al., 2007) that excels at learning interpretable representations of human physiology (Dupré la Tour et al., 2018; Power et al., 2023; Chen et al., 2023): many human activities are structured and tend to exhibit periodic patterns, which can be captured by CDL. During locomotion for instance—our illustrative example throughout this text—CDL provides a principled way to statistically infer the gait cycles directly from the data (Whittle, 2014), thereby circumventing the need for experts to craft a dictionary themselves (typically for template matching, *cf.*, Ying et al., 2007; Micó-Amigo et al., 2016; Oudre et al., 2018; Dot et al., 2020; Voisard et al., 2024).

Nonetheless, vanilla CDL is unable to handle a key characteristic of human physiological data, namely their *local-global* structure. At the population-level, *global*, shared, patterns may emerge due to similar anatomical traits or common neuronal circuits. At the individual’s scale, *local*, personal variations around these commonalities may exist due to various idiosyncrasies, differences in perceptual-motor style or pathology (Mantilla et al., 2020; Vidal and Lachauquini, 2021; Webster and Darter, 2019). Naive application of CDL to this type of data (hereafter referred to as PopCDL for “population-level CDL”) will yield a globally shared *population dictionary* that disregards individual-level structure, treating it as noise. Conversely, applying CDL to each individual independently (hereafter referred to as IndCDL for “individual-level CDL”) precludes the learning of the

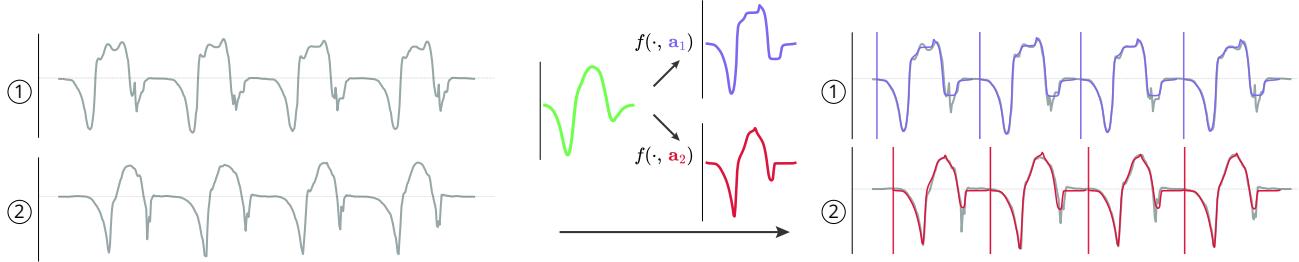


Figure 1: **Illustration of PerCDL for the Analysis of Human Locomotion Data.** (left) The foot kinematic signals of two individuals ① and ② present a repeated structure called a gait cycle. (middle) PerCDL learns the shared structures and subject-specific variability: the common shape (green) is “personalized” using a transformation function f (parameterized by the personalization parameters \mathbf{a}_1 and \mathbf{a}_2). (right) PerCDL successfully identifies all gait cycles and signal-specific shapes (blue and red), resulting in an accurate reconstruction.

shared structures in the data, and can be unreliable under the presence of noise. Despite efforts to address these challenges, the global-local structure has not been fully leveraged to achieve meaningful representations of human physiology (see Section 4.1).

To address these issues, we propose Personalized CDL (PerCDL), which aims to infer both global (population-level structures) and local (individual-level peculiarities) patterns from a collection of time series. The reconstruction process of PerCDL is illustrated in Figure 1. Briefly, PerCDL learns individual-specific personalized atoms from a set of common atoms shared across the population. Personalization is achieved through spatiotemporal transformation of the common atoms using Time Warping (TW; Müller, 2007). Specifically, we adapt the shape reparameterization method of Celledoni et al. (2023) to the CDL context. Unlike previous works that also investigated the use of TW in DL (Xu et al., 2023; Yazdi and Douzal-Chouakria, 2018), our approach does not require the manual selection of a set of increasing functions (Xu et al., 2023) and is computationally more efficient than that of Yazdi and Douzal-Chouakria (2018), making it more robust, scalable, and usable.

While the proposed framework is general in nature, particular consideration will be given to the analysis of neurophysiological signals (*e.g.*, heart rate, breathing, walking, etc.). Specifically, the analysis of gait. To better model such signals, a common practice in CDL has been to impose non-overlapping atoms to reflect the underlying physiological origin of these events (German et al., 2024; Zhu et al., 2016; Schäfer and Leser, 2022; Torkamani and Lohweg, 2017; Jas et al., 2017; Dupré la Tour et al., 2018; Power et al., 2023). The no-overlap assumption will be implicit in the remainder of this text. Besides being physiologically relevant, it is also computationally convenient enabling closed-form updates of the common dictionary (the *exact* solution of the l_0 sparse coding problem; Truong and

Moreau, 2024) and parallelism at each step. Furthermore, this constraint allows us to conduct an analysis of the convergence rate of the associated Maximum Likelihood Estimator (MLE). For applications where the no-overlap assumption is no longer realistic, any alternative optimisation method that relaxes this assumption (Song et al., 2018; Toloochams et al., 2020) can be used instead.

- We propose PerCDL¹, a framework for learning interpretable representations of time-series data that highlight personal deviations from common structures (Section 2.2).
- We present a meta-algorithm for solving the PerCDL problem (Section 2.3). The algorithm is modular in design, offering the flexibility to combine various components from conventional CDL (Garcia-Cardona and Wohlberg, 2018; Truong and Moreau, 2024). A federated learning, fully parallelizable variant specific to physiological time series representation is also introduced (Appendix B). Its complexity is $\mathcal{O}(N \log(N))$ per processor, where N is the number of samples per time series.
- We provide statistical guarantees for PerCDL (Section 2.5, Theorem 1): the estimated dictionary converges towards the true common dictionary at a rate of $\mathcal{O}(1/\sqrt{Sp})$, with S the number of individuals and p the number of pattern observations per individual. This improves over the naive scheme, IndCDL, which achieves a rate of $\mathcal{O}(1/\sqrt{p})$.
- We evaluate PerCDL on synthetic and real-world human physiology data, namely locomotion and ECG signals (Section 3). The results demonstrate that PerCDL reliably extracts interpretable global and local motifs while being robust to noise.

2 PERSONALIZED CDL

Notations. The integer range $\{k, \dots, l\} \subset \mathcal{P}(\mathbb{Z})$ is written $[k : l]$, and by extension $[l] = [1 : l]$, with

¹<https://github.com/axelroques/PerCDL>.

```

1 Inputs  $\mathbf{X} \in \mathbb{R}^{S \times N \times P}$ ,  $f : (\mathbb{R}^P)^L \times \mathbb{R}^M \mapsto (\mathbb{R}^P)^L$ 
2 Outputs  $\Phi \in \mathbb{R}^{K \times L \times P}$ ,  $\mathbf{A} \in \mathbb{R}^{S \times K \times M}$ ,  $\mathbf{Z} \in \mathbb{R}^{S \times K \times N-L+1}$ 
1  $\Phi, \mathbf{A}, \mathbf{Z} \leftarrow \text{setInitialValues}()$  // Initialization
// Obtain initial estimate for the common dictionary
2 while  $i \leq n_{\text{init}}$  do
3    $\mathbf{Z} \leftarrow \text{CSC}(\mathbf{X}, \mathbf{Z}, \Phi)$ : for any  $s \in [S]$ ,  $\underset{\mathbf{Z}_s \geq 0}{\text{argmin}} \left\| \mathbf{x}^s - \sum_{k=1}^K \mathbf{z}_k^s * \phi_k \right\|_2^2 + \lambda \|\mathbf{Z}\|_0$ ;
4    $\Phi \leftarrow \text{CDU}(\mathbf{X}, \mathbf{Z}, \Phi)$ :  $\underset{\Phi: \|\phi\|_2=1}{\text{argmin}} \sum_{s=1}^S \left\| \mathbf{x}^s - \sum_{k=1}^K \mathbf{z}_k^s * \phi_k \right\|_2^2$ ;
5 end
// Personalization
6 while  $i \leq n_{\text{perso}}$  do
7    $\mathbf{A} \leftarrow \text{IPU}(\mathbf{X}, \mathbf{Z}, \Phi, \mathbf{A}, f)$ : for any  $s \in [S]$ ,  $\underset{\mathbf{A}}{\text{argmin}} \left\| \mathbf{x}^s - \sum_{k=1}^K \mathbf{z}_k^s * f(\phi_k, \mathbf{a}_k^s) \right\|_2^2$ ;
8    $\mathbf{Z} \leftarrow \text{CSC}(\mathbf{X}, \mathbf{Z}, \hat{\Phi})$ ;
9    $\Phi \leftarrow \text{PerCDU}(\mathbf{X}, \mathbf{Z}, \Phi, \mathbf{A}, f)$ :  $\underset{\Phi: \|\phi_k\|_2=1}{\text{argmin}} \sum_{s=1}^S \left\| \mathbf{x}^s - \sum_{k=1}^K \mathbf{z}_k^s * f(\phi_k, \mathbf{a}_k^s) \right\|_2^2$ ;
10 end

```

Algorithm 1: Meta-algorithm to Solve the PerCDL Problem in Equation (3). $\hat{\Phi}$ is the personalized dictionary obtained from the transformation of the atoms in the common dictionary Φ .

$k, l \in \mathbb{N}$. For any vector $\mathbf{u} = (u_t)_{t \in [T]}$, we define its l_2 -norm by $\|\mathbf{u}\|_2 = \sqrt{\sum_{t=1}^T \|u_t\|^2}$ and its l_0 -norm by $\|\mathbf{u}\|_0 = \sum_{t=1}^T 1_{\neq 0}(u_t)$. We denote by $\mathcal{O}_P = \{O \in \mathbb{R}^{P \times P} : O^\top O = I_P\}$ and $C^1([0, 1])$ the space of continuously differentiable functions $g : [0, 1] \rightarrow [0, 1]$.

2.1 Convolutional Dictionary Learning

Consider a time-series dataset $\mathbf{X} \in \mathbb{R}^{S \times N \times P}$ of S P -dimensional time series $(\mathbf{x}^s)_{s \in [S]}$ of equal size N . In the context of human physiological signals, S may correspond to the number of unique individuals, P may refer to the number of sensor features, and N to the maximum number of samples per individual. Given a shared dictionary $\Phi = (\phi_k)_{k \in [K]} \in \mathbb{R}^{K \times L \times P}$ of K common patterns of length L ($L < N$), their respective weights, and activations $\mathbf{Z} = (\mathbf{z}_k^s)_{s, k \in [S] \times [K]}$ in $\mathbb{R}^{S \times K \times (N-L+1)}$, the CDL problem is stated as:

$$\underset{\mathbf{Z} \geq 0, \Phi: \|\phi_k\|_2=1}{\text{argmin}} \sum_{s=1}^S \left\| \mathbf{x}^s - \sum_{k=1}^K \mathbf{z}_k^s * \phi_k \right\|_2^2 + \lambda \|\mathbf{Z}\|_0, \quad (1)$$

where $*$ denotes the multivariate linear convolution such that, for two time series $\mathbf{z} = (z_i)_{i \in [N-L+1]} \in \mathbb{R}^{N-L+1}$ and $\mathbf{y} = (y_j)_{j \in [L]} \in (\mathbb{R}^P)^L$,

$$\mathbf{z} * \mathbf{y} = \left(\sum_{l=1}^L y_l z_{m-l+1} \right)_{m \in [N]} \in (\mathbb{R}^P)^N.$$

λ is a penalty term that balances the sparsity of the activation matrix and the reconstruction error. Increasing λ results in sparser activations and yields more interpretable atoms but often at the cost of an increased reconstruction error.

This problem formulation does not explicitly consider the local-global structure of physiological signals: one can either learn a dictionary for each individual by running CDL S -times (IndCDL) or, alternatively, learn a single dictionary that is shared across the whole population (PopCDL).

2.2 Personalized Convolutional Dictionary Learning

To accommodate inter-individual variability within the CDL framework, we propose to learn population-level atoms along with individual-level atoms. Each personalized atom $\hat{\phi}_k^s$ is derived from a common atom ϕ_k in the global dictionary Φ via a transformation function. More formally, denoting a parameterized transformation as $f : (\mathbb{R}^P)^L \times \Theta \rightarrow (\mathbb{R}^P)^L$, where $\Theta \subset \mathbb{R}^M$ is the set of personalization parameters, the personalized version of the k^{th} global atom ϕ_k for the s^{th} time series is expressed as, for any $(s, k) \in [S] \times [K]$,

$$\hat{\phi}_k^s = f(\phi_k, \mathbf{a}_k^s) \quad (2)$$

with $\mathbf{a}_k^s \in \Theta$, the personalization parameters. This representation strictly generalizes both IndCDL and PopCDL since they can be obtained by setting $f(\phi_k, \mathbf{a}_k^s) = \mathbf{a}_k^s \in (\mathbb{R}^d)^L$ or $f(\phi_k, \mathbf{a}_k^s) = \phi_k$ respectively. The choice of a suitable transformation function f is discussed in Section 2.4.

With $\mathbf{A} = (\mathbf{a}_k^s)_{s, k \in [S] \times [K]} \in \mathbb{R}^{S \times K \times M}$, the matrix of all personalization parameters, PerCDL solves

$$\underset{\mathbf{Z} \geq 0, \Phi: \|\phi_k\|_2=1, \mathbf{A}}{\text{argmin}} \sum_{s=1}^S \left\| \mathbf{x}^s - \sum_{k=1}^K \mathbf{z}_k^s * \hat{\phi}_k^s \right\|_2^2 + \lambda \|\mathbf{Z}\|_0. \quad (3)$$

In brief, the common structures of a dataset \mathbf{X} are identified in a dictionary Φ , while time series-specific variations of these global features are captured through the parameters matrix \mathbf{A} . Information about the occurrence and magnitude of these personalized motifs are contained in the activation matrix \mathbf{Z} . Figure 1 illustrates the principles of PerCDL.

2.3 A Meta-Algorithm

This section introduces a general meta-algorithm for solving the PerCDL problem in Equation (3) (Algorithm 1). The algorithm comprises two steps: ① Obtain an initial estimate of the common dictionary Φ ; ② Refine Φ to yield the personalized dictionary $\tilde{\Phi}$.

① Initial Estimates for Φ and \mathbf{Z} (Lines 2-5).

The classical CDL problem (Equation (1)) is first solved to obtain initial estimates for the common dictionary and the associated activations. Any off-the-shelf CDL algorithm can be used, such as a block-coordinate descent schemes, which alternate between a convolutional sparse coding step (CSC) and a convolutional dictionary update step (CDU). Given a fixed dictionary, the CSC step finds the best activations \mathbf{Z} subject to a sparsity-promoting penalization. In this work, the l_0 -penalty method proposed in Truong and Moreau (2024) is adopted as it is both robust and exact for non-overlapping atoms. Other approximate l_1 -penalized methods that relax this assumption could be used (Kavukcuoglu et al., 2010; Boyd et al., 2011; Moreau et al., 2018). Given a fixed set of activations \mathbf{Z} , the CDU step finds the dictionary Φ that minimizes the l_2 -reconstruction error. Once again, any dictionary update scheme can be used (Garcia-Cardona and Wohlberg, 2018).

② Personalization (Lines 6-10). Once sensible initial estimates for Φ and \mathbf{Z} have been found, the personalization parameters \mathbf{A} are learned (Line 7), while the activations \mathbf{Z} (Line 8) and the common dictionary Φ (Line 9) are refined in a block-coordinate descent scheme. In the individual parameters update (IPU) step, any constrained optimization algorithm can be used to obtain an approximate solution, such as projected gradient descent (Bubeck et al., 2015). This work uses gradient descent with the Polyak step-size (Hazan and Kakade, 2019), where the infimum of the objective is substituted with its lower bound, 0 (Loizou et al., 2021). This can be run in parallel on S processors. When updating the dictionary given \mathbf{A} and \mathbf{Z} (PerCDU step), the transformation f may introduce non-linearities that prohibit the use of conventional CDU update schemes, a problem shared with parametric DL methods (Ataee et al., 2010; Turquais et al., 2018; Chen et al., 2023). In fortunate cases

where f is linear in ϕ , conventional CDU updates can be exploited.

Federated Learning Implementation. Under the no-overlap assumption, the CDU step reduces to an average and can be computed efficiently as $\phi_k = \sum_j^{p_s^k} \mathbf{y}_k^s[j] / p_s^k$, denoting by $\mathbf{y}_k^s[j]$ the j^{th} subpart of \mathbf{x}^s where a pattern is recognized (a non-zero activation in \mathbf{z}_k^s) and p_s^k the number of non-zero activations in \mathbf{z}_k^s . Furthermore, a closed formed solution is also available for the PerCDU step (Theorem 2). These observations support the implementation of a federated learning version of this meta-algorithm, especially well-suited to the case of physiological data because of the aforementioned constraint (additional details can be found in Appendix B).

Initialization. Both the activations \mathbf{Z} and the personalization parameters \mathbf{A} are initialized to zero. This is common practice for \mathbf{Z} and empirical evidence suggest that this results in good performance for A . In contrast, it is known that DL and CDL are sensitive to dictionary initialization (Garcia-Cardona and Wohlberg, 2018; Agarwal et al., 2014; Ravishankar et al., 2020). In the context of overcomplete DL, Arora et al. (2013) and Agarwal et al. (2014) have provided upper bounds on the distance between the initial estimate and the true underlying dictionary to ensure complete recovery, as well as methods to achieve such theoretically good initializations. Therefore, a dictionary initialization with a reasonable estimate of the expected patterns is recommended whenever possible. In our experiments, the common atoms in Φ were initialized with a representative motif of the expected common structures (e.g., a complete gait cycle obtained from a healthy participant).

Time Complexity. Given that a basic step of CDL is an alternating minimization scheme (CDU+CSC), the complexity of N_{step} of CDL is thus $\mathcal{C}[N_{\text{step}}](CDL) = N_{\text{step}}[\mathcal{C}(\text{CSC}) + \mathcal{C}(\text{CDU})]$, with $\mathcal{C}(\text{CSC}) = \mathcal{O}(SKLN(\log(N) + P))$ (Truong and Moreau, 2024) and $\mathcal{C}(\text{CDU}) = \mathcal{O}(SNKP)$ (Garcia-Cardona and Wohlberg, 2018). If the transformation f is linear in ϕ , $\mathcal{C}(\text{PerCDU}) = \mathcal{C}(\text{CDU})$. Therefore, the complexity of PerCDL can be written as $\mathcal{O}(\mathcal{C}[n_{\text{init}} + n_{\text{perso}}](CDL) + n_{\text{perso}}\mathcal{C}(\text{IPU}))$. The extra cost of PerCDL compared to PopCDL is only the parameter update (IPU), which can be computed in parallel. $\mathcal{C}(\text{IPU})$ is bounded by $\mathcal{O}(SKMNPT_{\text{grad}})$ with T_{grad} the number of gradient steps. However, Lemma 2 on orthogonal transformation (Appendix A.1) proves that the IPU can be solved more efficiently if structural assumptions can be made on f .

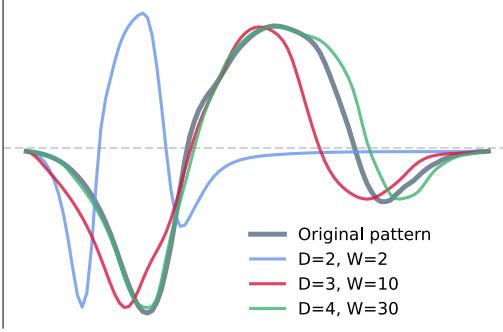


Figure 2: Time Warping using Eqs. (4)-(5).

2.4 Transformation Function

The choice of the transformation function f is paramount to the PerCDL framework and should be regarded as an integral aspect of downstream problem modelling. Since our framework is general, any transformation function appropriate for the task at hand can be used. This section will provide a family of transformation functions that are well suited for, but not exclusive to, human physiology data.

Time Warping Transformation. Let us consider a continuous signal $s : [0, 1] \rightarrow \mathbb{R}^P$. For $t \in [0, 1]$, the “time-warped” version of s is $s_{\text{warp}}(t) = s(\psi_{\mathbf{a}}(t))$, where $\psi_{\mathbf{a}} : [0, 1] \rightarrow [0, 1]$ is a time warping function parameterized by $\mathbf{a} \in \mathbb{R}^M$. In our context, a time-warp $\psi_{\mathbf{a}} : [0, 1] \rightarrow [0, 1]$ is applied to each atom $\phi = (\phi_l)_{l \in [L]} \in (\mathbb{R}^P)^L$ (interpreted as a discrete-time signal) as, for any $i \in [L]$,

$$f(\phi, \mathbf{a}) = \mathcal{T}_{\psi_{\mathbf{a}}, \sigma}(\phi), \quad [\mathcal{T}_{\psi_{\mathbf{a}}, \sigma}(\phi)]_i = \tilde{\phi}(\psi_{\mathbf{a}}(t_i)), \quad (4)$$

where $t_i = i/L$ is the i th sampling time, $\tilde{\phi} : [0, 1] \mapsto (\mathbb{R}^P)^L$ is an interpolation of ϕ , and $\mathcal{T}_{\psi_{\mathbf{a}}, \sigma} : (\mathbb{R}^P)^L \rightarrow (\mathbb{R}^P)^L$ is the composition of the interpolation of ϕ and the time warping $\psi_{\mathbf{a}}$.

A smooth linear interpolation is used to interpolate ϕ . Let $c_l(t) = (t - t_l)$ for any $(t, l) \in [0, 1] \times [L]$. Then,

$$\tilde{\phi}(t) = \sum_{l=1}^L w_l(\sigma, t) [\phi_l + c_l(t)(\phi_{l+1} - \phi_l)],$$

where $w_l(\sigma, t) = \exp\left[-\frac{c_l(t)^2}{2\sigma^2}\right] / \sum_{l'=1}^L \exp\left[-\frac{c_{l'}(t)^2}{2\sigma^2}\right]$, and σ is a regularization parameter such that $w_l(\sigma, t) \rightarrow \mathbb{1}_{(l-\frac{1}{2}, l+\frac{1}{2})}(Lt)$ as σ goes to zero. Conveniently, the resulting transformation f is linear in ϕ . Furthermore, w_l allows for a smooth, differentiable mapping $(\phi, \mathbf{a}) \mapsto f(\phi, \mathbf{a})$ so that gradient-based optimization can be leveraged for the IPU update.

Time Warping Function $\psi_{\mathbf{a}}$. We opt for the time warping parameterization presented by Celledoni et al.

(2023), which takes advantage of the Lie group structure of

$$\mathcal{D}_{\psi} = \{g \in C^1([0, 1]) : g(0) = 0, g(1) = 1, g' > 0\},$$

the set of non-degenerate time warpings. (Degenerate time warpings allow $g' \geq 0$ such that g has not necessarily an inverse.) \mathcal{D}_{ψ} is a group for the composition: the composition of two time warps is a valid time warp. Thus, an infinitesimal displacement around $g \in \mathcal{D}_{\psi}$ is not $g + dg$, but $(Id + dg) \circ g$, where dg is in $\text{TM}_g = \{g \in C^1([0, 1]) : g(0) = g(1) = 0\}$, the tangent space at g (Celledoni et al., 2023). Given two hyper-parameters $D, W \in \mathbb{N}$ and $M = DW$, we define, for any $\mathbf{a} = (a_w^d)_{d, w \in [D] \times [W]} \in \Theta$, $t \in [0, 1]$, and $d \in [D]$, $\psi_{\mathbf{a}} = \psi_{D, \mathbf{a}} \circ \dots \circ \psi_{1, \mathbf{a}}$ as a composition of D displacements:

$$\psi_{d, \mathbf{a}} = \text{Id} + \sum_{w=1}^W a_w^d b_w, \quad b_w(t) = \frac{\sin(w\pi t)}{w\pi}. \quad (5)$$

In the following, we motivate the choice of Θ such that $\psi_{\mathbf{a}}$ is in the set of time warping \mathcal{D}_{ψ} , i.e., $\psi_{\mathbf{a}}[0] = 0$, $\psi_{\mathbf{a}}[1] = 1$ and $\psi'_{\mathbf{a}} > 0$. Note that $(b_w)_{w \geq 1}$ is a Fourier basis of TM_g , thus by construction $\psi_{\mathbf{a}}[0] = 0$, $\psi_{\mathbf{a}}[1] = 1$. Moreover, b_w is normalized such that $|b'_w| \leq 1$. As such, if for any $d \in [D]$, $\sum_{w=1}^W |a_w^d| < 1$, then we have $\psi'_{\mathbf{a}, d} \geq 0$ for any $d \in [D]$ and therefore by composition of time warping $\psi_{\mathbf{a}} \in \mathcal{D}_{\psi}$. Hence, the parameter set Θ is chosen as

$$\left\{ (a_w^d)_{d, w \in [D] \times [W]} \in \mathbb{R}^M : \sum_{w=1}^W |a_w^d| \leq 1, d \in [D] \right\}.$$

The depth D and width W of the transformation both shape the extent of the variability around the common atoms. Tuning these hyper-parameters offers a trade-off between flexibility and interpretability of the transformation. D controls the Lipschitz constant of ψ , i.e., *global dilation*, and W controls the high frequency truncation, i.e., *local dilation*.

Figure 2 shows the effect of $f(\cdot, \mathbf{a})$ defined with (4)-(5) for different choices of hyper-parameters D, W and uniformly sampled \mathbf{a} from $(-1, 1)^M$ projected onto Θ . Higher values of D/W allow for greater flexibility, while lower values increase the granularity of the transformation. As illustrated in Figure 1, this class of transformation is general enough to recover pathological gait cycles from a healthy one. Additional experiments on the sensitivity of the transformations with respect to D and W are presented in Appendix C.2 of the appendix.

2.5 Theoretical Guarantees: Toward Mixed-Effects Models

PerCDL and mixed-effects models are naturally related since they both combine population-level and

individual-level parameters. In mixed-effects models, high-level structures in the population characteristics have been shown to boost the convergence rate of the MLE under mild assumptions, to a rate of $O(1/\sqrt{Sp})$ for population parameters (Nie, 2007, Theorem 3), or $O(1/\sqrt{p})$ for individual parameters, with S the number of individuals and p the number of observations per individual. This section will prove a similar convergence rate of $O(1/\sqrt{Sp})$ for the MLE related to PerCDL's common atoms, under reasonable assumptions.

Theoretical Setup. For the sake of clarity, our interest will be limited to a single common pattern in $\Phi = \{\phi_*\}$ (*i.e.*, $K = 1$), where $\phi_* \in (\mathbb{R}^P)^L$ is the true common atom. This implies that, for any $s \in [S]$, denoting the s^{th} parameter as $a_*^s \in \mathbb{R}^M$, the observation is generated as

$$\mathbf{z}^s = \mathbf{z}^s * f(\phi_*, a_*^s) + \epsilon^s, \quad \epsilon^s \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0_N, \sigma^2 I_N). \quad (6)$$

Let us denote by $\mathbf{y}^s[j]$ the j^{th} subpart of \mathbf{z}^s where a pattern is recognized by a non-zero activation in \mathbf{z}^s . Now, provided that there is no overlap between atoms—a realistic assumption in physiological signals—and that \mathbf{z}^s is known—another weak assumption since l_0 penalization is used and its sparse-coding solver is exact under non-overlapping condition (Truong and Moreau, 2024)—for any $s, j \in [S] \times [p_s]$, $\mathbf{y}^s[j]$ can be decomposed as

$$\mathbf{y}^s[j] = f(\phi_*, a_*^s) + \epsilon^s[j], \quad \epsilon^s[j] \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0_L, \sigma^2 I_L), \quad (7)$$

where p_s is the total number of observations and $a_*^s \stackrel{\text{i.i.d.}}{\sim} \mathcal{P}$ the distribution of individual parameters.

Even under these seemingly strong assumptions, the analysis will prove to be non-trivial because the assumptions on the distribution of individual parameters \mathcal{P} are weak and the transformation function may be highly non-linear (TW). This work provides the first theoretical analysis of DL with TW (*cf.*, Yazdi and Douzal-Chouakria, 2018; Xu et al., 2023). Identifying the minimal set of necessary assumptions on \mathcal{P} is particularly important to gain insight into the group structure necessary for the recovery of common atoms.

Case 1: Global structure is absent. If there is no shared, global structure in the population (as assumed by IndCDL, *i.e.*, $f(\phi_*, a_*^s) = a_*^s \in (\mathbb{R}^P)^L$), then the MLE estimator is the local empirical mean $\frac{1}{p_s} \sum_{j=1}^{p_s} \mathbf{y}^s[j]$ and converges at a rate of $O(1/\sqrt{p_s})$.

Case 2: Global structure is present. In this case, we consider the following hypothesis to derive a convergence rate for the MLE on the common atom ϕ_* .

H 1. *There exists a common atom $\phi_* \in (\mathbb{R}^P)^L$ with individual parameters $a_*^s \in \mathbb{R}^M$, and an operator*

$\mathcal{L} : \mathbb{R}^M \rightarrow \mathcal{M}_{L,L}(\mathbb{R})$, such that for any $s \in [S]$, $\phi_^s = \mathcal{L}(a_*^s) \phi_*$. Moreover, \mathcal{P} is a distribution on \mathbb{R}^M such that $\mathbb{E}_{\mathcal{P}}(|\mathcal{L}(a_*^s)^\top \mathcal{L}(a_*^s)|) < \infty$ and $\mathbb{E}_{\mathcal{P}}(\mathcal{L}(a_*^s)^\top \mathcal{L}(a_*^s))$ is symmetric positive definite. Its minimum eigenvalue is denoted by $\rho > 0$.*

Under **H1**, the model is a linear mixed-effects model if and only if \mathcal{P} is Gaussian (Nie, 2007). In **H1**, \mathcal{P} needn't be Gaussian, *e.g.*, it can be a multimodal distribution. Therefore, **H1** can be seen as a refinement of Condition 2 by Nie (2007).

Lemma 1. *The assumption **H1** is met in the two following cases:*

- *Time warping transformations: For any $\mathbf{a} \in \Theta \subset \mathbb{R}^M$, $\mathcal{L}(\mathbf{a}) = \mathcal{T}_{\psi_{\mathbf{a}}, \sigma}$ where $\mathcal{T}_{\cdot, \sigma}$ and ψ are defined in (4) and (5) with σ small enough. \mathcal{P} is a distribution supported on Θ such that for any $l \in [L]$, the l -th sample of the common atom ϕ has a positive probability to be seen in a personalized atom ϕ_*^s .*
- *Rotations: For any $\mathbf{a} \in \Theta \subset \mathbb{R}^M$, $\mathcal{L}(\mathbf{a}) \in \mathcal{O}_P(\mathbb{R})$.*

More formal statements and proofs are postponed to Appendix A.

The following additional hypothesis is needed to invoke a weighted law of large numbers in Theorem 1 as it prevents an imbalance between the individuals' observations.

H 2. *There exists a constant $C < +\infty$ such that for any $s \in [S]$, $1 \leq p_s \leq C$.*

Taking advantage of the fact that $\beta \in (\mathbb{R}^P)^L \rightarrow f(\mathbf{a}, \beta)$ is linear, we derive the following theorem.

Theorem 1. *Under **H1**, **H2**, and assuming that the true parameters $(a_*^s)_{s \in [S]}$ are given, for S large enough, $\Sigma_{S,p} = \sum_{s=1}^S p_s \mathcal{L}(a_*^s)^\top \mathcal{L}(a_*^s)$ is invertible and the MLE estimator $\hat{\phi}_{\text{mle}} \sim \mathcal{N}(\phi_*, (\Sigma_{S,p})^{-1})$ is Gaussian. Moreover, $\hat{\phi}_{\text{mle}}$ converges toward ϕ_* at a rate of $\mathcal{O}(1/\rho \sqrt{\sum_{s=1}^S p_s})$ in probability.*

When $p_s = p$ for any $s \in [S]$, we obtain the same convergence rate as for population parameters in mixed-effect models $\mathcal{O}(1/\sqrt{Sp})$ (Nie, 2007, Theorem 3). Furthermore, the assumption that the true parameters $(a_*^s)_{s \in [S]}$ are known can be relaxed by assuming that a \sqrt{p} -consistent estimator $(\hat{a}_s^*)_{s \in [S]}$ of $(a_*^s)_{s \in [S]}$ is available. However, more technical conditions (Nie, 2007, Condition 1) would have to be involved and complicate the exposition. Nevertheless, Theorem 1 asserts that, under suitable conditions, PerCDL offers identical statistical guarantees as mixed-effects models for estimating the common parameters.

3 EXPERIMENTS

The meta-algorithm presented in Section 2.3 was applied to synthetic (Section 3.1) and real-world datasets (locomotion data in Section 3.2 and ECG data in Appendix C.6). When applicable, PerCDL was compared with IndCDL and PopCDL. Implementation details and additional numerical experiments can be found in Appendix C.

3.1 Synthetic Data

We generate a synthetic dataset with S time series of equal length N . Each time series contains $r = 3$ repetitions of $K = 2$ common patterns of length L (Figure 7). Both patterns are personalized independently for each signal through TW (Equation (4)), with parameters drawn uniformly random from $(-1, 1)^M$ and projected onto Θ . The signal-specific patterns are then placed at randomly chosen positions within each time series, with no overlap. In all experiments, activations and personalization parameters are initialized to zero, while the dictionary is initialized with the personalized atom of the first signal.

Identifiability. We investigate the ability of IndCDL, PopCDL, and PerCDL to extract the common structures with an increase in the number of signals. All methods are applied to multiple synthetic datasets that contain between $S = 1$ and $S = 1024$ time series of equal length $N = 500$. For IndCDL, the common structures are computed as the Euclidean barycenter of the individual atoms found in each signal. While we also considered a barycenter computed with the Soft-DTW metric, this led to worse results. Thus, we only consider the Euclidean barycenter to ensure a fair comparison against PopCDL and PerCDL. The distance between the obtained common structures and the underlying ground truths are shown in Figure 3. (The shape of the common atoms estimated by the three methods is shown in Figure 10 in the Appendix.)

As the number of signals in the dataset increases, the quality of the common atoms identified by IndCDL and PopCDL degrades. This suggests that the idiosyncrasies of each signal negatively affects estimation of the common structures. Notably, IndCDL appears to perform better than PopCDL as the number of signals increases. We suspect that the common structures identified by PopCDL, by construction, do not match the local variations of the individual signals induced by the time warps, which could affect the segmentation performance of the CSC step. This segmentation error may in turn lead to CDU updates that drift further and further away from the true underlying common structures when the amount of individual atom vari-

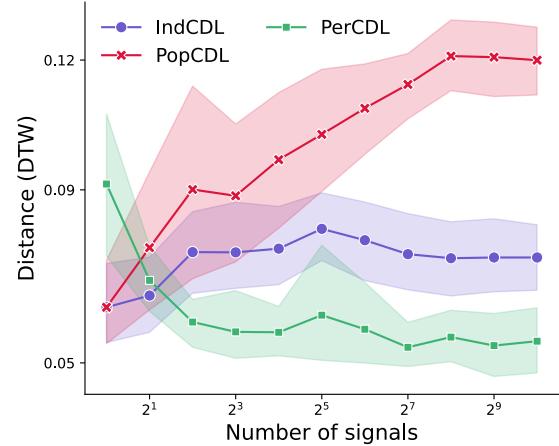


Figure 3: **Convergence Toward the Common Structures.** Average distance between the common atoms identified in the synthetic experiment and the ground truth as a function of the dataset’s size. Shaded regions represent the 95% confidence interval.

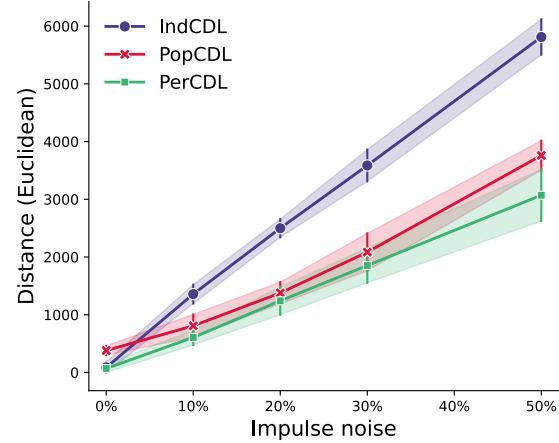


Figure 4: **Reconstruction Error Under Impulse Noise Contamination.** Error bars represent two standard deviations.

ability increases. With IndCDL, segmentation is necessarily accurate since it uses the personalized structures of each signal directly. In contrast with IndCDL and PopCDL, PerCDL ensures a good identifiability of the common shapes in the data that improves as the size of the dataset increases.

Robustness to Noise. We now evaluate the reconstruction accuracy of PerCDL under impulse noise. (Additional experiments with Gaussian noise can be found in Appendix C.4). To that end, we generate a synthetic dataset of $S = 32$ time series of equal length $N = 1000$. In each time series, $r = 3$ patterns of length $L = 50$ are placed at random positions, with no

overlap. A quarter of the signals are then corrupted with impulse noise: p percent of all samples—chosen at random—are polluted with either a positive or negative spike of uniformly random amplitude in the range [2, 2.5]. The reconstruction errors as a function of the proportion of corrupted samples are presented in Figure 4.

While the performance of all methods degrades with an increase in noise, PerCDL proves more robust than IndCDL and PopCDL. Our results indicate that population-level structure information combined with the superior representation possibilities offered by the personalization step better withstand noise and artefacts in the input data.

3.2 Application to Gait Analysis

We now apply PerCDL to real-world healthy and pathological locomotion data (Truong et al., 2019) (more details are provided in Appendix C.5). 150 participants were instructed to walk at their preferred speed in a straight line on a leveled surface for 10 meters, while an inertial measurement unit measured their foot angular velocity in the sagittal plane. Participants were divided into three equal-sized groups of size $n = 50$ (Healthy, Orthopedic, and Neurological) according to their health status. The common dictionary was initialized with the gait cycle extracted from a randomly chosen healthy participant.

One of the first task in gait analysis is to identify the gait cycles within each locomotion signal. PerCDL’s ability to successfully recover the gait cycles, through the activation matrix Z , is summarized in Table 1. We obtain results comparable to the state-of-the-art gait segmentation method introduced in Voisard et al. (2024). More details on this experiment can be found in Appendix C.5.

Table 1: **Segmentation Performance of PerCDL vs. SOTA.** FPR: False Positive Rate.

| | Voisard et al. | PerCDL |
|--------------------|-----------------------|---------------|
| Sensitivity | 0.959 | 0.988 |
| FPR | 0.004 | 0.033 |

However, PerCDL is not only capable to efficiently segment locomotion signals, it can also be used to represent, interpret and analyze such data. Figure 5(a) shows that the common atom identified (grey line) is similar to the average of the personalized atoms in the healthy population. Intuitively, the shared characteristics within a population should roughly resemble the average of its healthy constituents. Thus, the shared structure identified by PerCDL carries an intrinsic meaning and serves as a coherent foundation

for signal-specific personalization through time warps. This finding is promising as it could facilitate the monitoring of fine deviations from a “healthy baseline” that may be indicative of the start of a pathological process (König et al., 2016; Vidal and Lacquaniti, 2021).

We then examine the personalized atoms found by PerCDL in the different pathology groups (Figure 14). While all groups present similar gait cycles, discernible inter-individual differences can be identified in the foot kinematics during specific phases of the gait cycle. Figure 5(b) highlights this point via the representation of the normalized relative difference in local gait variability between each pathological group and the healthy population. Our results suggest that the Orthopedic group diverges from a healthy gait cycle mainly at the beginning of the swing phase (“Toes off”), while the neurological group differs mainly at the end of the stance phase (between “Heel off” and “Toes off”) and right before the end of the swing phase (“Heel on”).

More experiments, including a comparison between IndCDL, PopCDL and PerCDL, are available in the Appendix (Appendix C.5). Overall, our findings demonstrate the effectiveness of PerCDL for gait analysis in the identification of both common and personal structures.

Furthermore, experiments on an ECG dataset in Appendix C.6 reveals that personalization parameters extract relevant characteristics from the input signals, such that they can be used as features in pathology classification.

4 DISCUSSION

4.1 Related Works

CDL (Grosse et al., 2007; Zheng et al., 2016; Mørup et al., 2008) is a well-established method for time series representation that incorporates translation-invariance to the DL framework (Kreutz-Delgado et al., 2003; Tošić and Frossard, 2011). Interestingly, CDL shares common roots with Convulsive Non-Negative Matrix Factorization (cNMF or cNMFsc, the latter imposing sparseness constraints; Smaragdis, 2006; O’Grady and Pearlmutter, 2008; Wang et al., 2009), which has been employed in speech separation and the analysis of articulatory movement primitives during speech production (Ramanarayanan et al., 2013). Specifically, Ramanarayanan et al. (2013); Smaragdis (2006) have analyzed individual speech patterns by examining their constituent atoms using a method similar to IndCDL. Theoretical connections can also be made with the deep learning litterature: PopCDL may be understood as a very specific Vec-

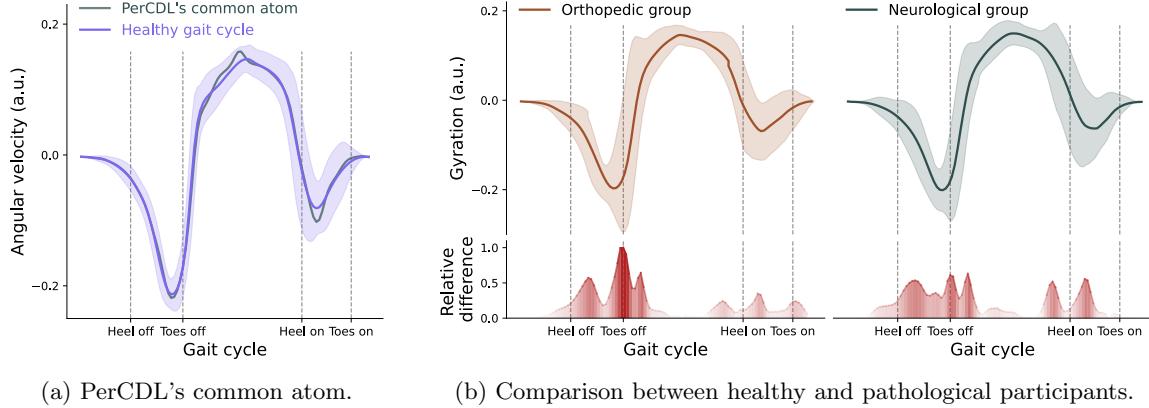


Figure 5: Application of PerCDL to Locomotion Data. (a) Common atom identified by PerCDL (grey) and averaged personalized gait cycle of the healthy population (blue). The shaded region represents two standard deviations. (b) Population-averaged personalized atoms learned by PerCDL in the Orthopedic (brown) and Neurological (green) groups, with two standard deviations (shaded areas). The red-colored curves represent the normalized relative variability in the gait cycle for the current population compared with the Healthy group.

tor Quantized Variational Auto-Encoder (VQ-VAE) model (Van Den Oord et al., 2017) if the common dictionary $\Phi \in \mathbb{R}^{K \times L \times P}$ is interpreted as the discrete latent space. Nonetheless, the projection step onto the closest vector is performed in the latent space in VQ-VAE but occurs during the sparse-coding step in CDL. Consequently, after the sparse coding step, we only need to learn the discrete embeddings (the atoms) from the observations.

Multiple extensions of CDL were developed to further enforce structure into the atoms. For instance, Soh proposed group-invariant DL, which complements the shift-invariance property with rotation invariance. Others have introduced parametric forms on the atoms to analyze seismic data (Chen et al., 2023; Turquais et al., 2018) or sketch vectorization (Shaheen et al., 2017). Parameterisation of the atoms is also central to PerCDL, though specifically to model global and local structures.

In parallel, the advent of individual-level data has precipitated the development of methodologies designed to create personalised representations in a plethora of applications from federated learning and unsupervised learning (Dinh et al., 2020; Huang et al., 2023). In representation learning in particular, personalised variants of principle component analysis (Ozkara et al., 2023; Shi and Kontar, 2024), matrix/tensor decomposition (Hu et al., 2023; Shi et al., 2023), and dictionary learning (Liang et al., 2024) were introduced. Although our work follows the aforementioned trend by proposing a personalized variant of CDL, it differs in that it considers *personalizations* that are specific to time-series data. The most closely related approach is perhaps the Generalized Canonical Time Warping

method (GCTW; Xu et al., 2023). GCTW was proposed as a means of resolving temporal misalignment between the input data and the atoms by jointly learning the dictionary and time warping matrices. Unlike PerCDL, however, GCTW is a DL method that focuses on *counteracting* temporal variability and therefore does not attempt to learn an interpretable representation of the local variations.

4.2 Conclusion

This work introduced Personalized Convolutional Dictionary Learning (PerCDL), a framework for learning interpretable representations of time-series datasets that captures both global structures and local variations around these commonalities. A meta-algorithm was proposed and its theoretical performance was evaluated. Theoretical consistency guarantees were obtained under the typical conditions used in the mixed-effects models literature. The performance of PerCDL was empirically demonstrated on synthetic and real-world human physiological signals.

The proposed framework is general in nature and could be extended to other time-series-related tasks by adapting the transformation function. The performance of PerCDL on other modalities, (such as images; (Shaheen et al., 2017)), is an interesting avenue for research and is left to future works.

The primary limitation of the proposed method is that of computational cost associated with the IPU step. To alleviate any potential scalability issue, a federated-learning variant of PerCDL was proposed. This represents a promising research direction that will be further refined in subsequent work.

Acknowledgments

K. Kim was supported by a grant from AWS AI to Penn Engineering’s ASSET Center for Trustworthy AI.

References

- Kenneth Kreutz-Delgado, Joseph F Murray, Bhaskar D Rao, Kjersti Engan, Te-Won Lee, and Terrence J Sejnowski. Dictionary learning algorithms for sparse representation. *Neural computation*, 15(2):349–396, 2003.
- Ivana Tošić and Pascal Frossard. Dictionary learning. *IEEE Signal Processing Magazine*, 28(2):27–38, 2011.
- Cristina Garcia-Cardona and Brendt Wohlberg. Convolutional dictionary learning: A comparative review and new algorithms. *IEEE Transactions on Computational Imaging*, 4(3):366–381, 2018.
- Roger Grosse, Rajat Raina, Helen Kwong, and Andrew Y Ng. Shift-invariant sparse coding for audio classification. In *Proceedings of the Twenty-Third Conference on Uncertainty in Artificial Intelligence*, pages 149–158, 2007.
- Tom Dupré la Tour, Thomas Moreau, Mainak Jas, and Alexandre Gramfort. Multivariate convolutional sparse coding for electromagnetic brain signals. *Advances in Neural Information Processing Systems*, 31, 2018.
- Lindsey Power, Cédric Allain, Thomas Moreau, Alexandre Gramfort, and Timothy Bardouille. Using convolutional dictionary learning to detect task-related neuromagnetic transients and ageing trends in a large open-access dataset. *NeuroImage*, 267: 119809, 2023.
- Hongling Chen, Mauricio D Sacchi, and Jinghuai Gao. Parametric convolutional dictionary learning and its applications to seismic data processing. *IEEE Transactions on Geoscience and Remote Sensing*, 2023.
- Michael W Whittle. *Gait analysis: an introduction*. Butterworth-Heinemann, 2014.
- Hong Ying, Carmen Silex, Andreas Schnitzer, Steffen Leonhardt, and Michael Schiek. Automatic step detection in the accelerometer signal. In *4th International Workshop on Wearable and Implantable Body Sensor Networks (BSN 2007) March 26–28, 2007 RWTH Aachen University, Germany*, pages 80–85. Springer, 2007.
- M Encarna Micó-Amigo, Idsart Kingma, Erik Ainsworth, Stefan Walgaard, Martijn Niessen, Rob C van Lummel, and Jaap H van Dieën. A novel accelerometry-based algorithm for the detection of step durations over short episodes of gait in healthy elderly. *Journal of neuroengineering and rehabilitation*, 13:1–12, 2016.
- Laurent Oudre, Rémi Barrois-Müller, Thomas Moreau, Charles Truong, Aliénor Vienne-Jumeau, Damien Ricard, Nicolas Vayatis, and Pierre-Paul Vidal. Template-based step detection with inertial measurement units. *Sensors*, 18(11):4033, 2018.
- Tristan Dot, Flavien Quijoux, Laurent Oudre, Aliénor Vienne-Jumeau, Albane Moreau, Pierre-Paul Vidal, and Damien Ricard. Non-linear template-based approach for the study of locomotion. *Sensors*, 20(7): 1939, 2020.
- Cyril Voisard, Nicolas De l’Escalopier, Damien Ricard, and Laurent Oudre. Automatic gait events detection with inertial measurement units: healthy subjects and moderate to severe impaired patients. *Journal of NeuroEngineering and Rehabilitation*, 2024.
- Juan Mantilla, Danping Wang, Ioannis Bargiotas, Junhong Wang, Jiuwen Cao, Laurent Oudre, and Pierre-Paul Vidal. Motor style at rest and during locomotion in humans. *Journal of Neurophysiology*, 123(6):2269–2284, 2020.
- Pierre-Paul Vidal and Francesco Lacquaniti. Perceptual-motor styles. *Experimental Brain Research*, 239(5):1359–1380, 2021.
- Joseph B Webster and Benjamin J Darter. Principles of normal and pathologic gait. In *Atlas of Orthoses and Assistive Devices*, pages 49–62. Elsevier, 2019.
- Meinard Müller. *Information retrieval for music and motion*, volume 2. Springer, 2007.
- Elena Celledoni, Helge Glöckner, Jørgen N Riseth, and Alexander Schmeding. Deep neural networks on diffeomorphism groups for optimal shape reparametrization. *BIT Numerical Mathematics*, 63 (4):50, 2023.
- Ruiyu Xu, Chao Wang, Yongxiang Li, and Jianguo Wu. Generalized time warping invariant dictionary learning for time series classification and clustering. *arXiv preprint arXiv:2306.17690*, 2023.
- Saeed Varasteh Yazdi and Ahlame Douzal-Chouakria. Time warp invariant ksvd: Sparse coding and dictionary learning for time series under time warp. *Pattern Recognition Letters*, 112:1–8, 2018.
- Thibaut Germain, Charles Truong, and Laurent Oudre. Persistence-based motif discovery in time series. *IEEE Transactions on Knowledge and Data Engineering*, 2024.
- Yan Zhu, Zachary Zimmerman, Nader Shakibay Senobari, Chin-Chia Michael Yeh, Gareth Funning, Abdullah Mueen, Philip Brisk, and Eamonn Keogh.

- Matrix profile ii: Exploiting a novel algorithm and gpus to break the one hundred million barrier for time series motifs and joins. In *2016 IEEE 16th international conference on data mining (ICDM)*, pages 739–748. IEEE, 2016.
- Patrick Schäfer and Ulf Leser. Motiflets: Simple and accurate detection of motifs in time series. *Proceedings of the VLDB Endowment*, 16(4):725–737, 2022.
- Sahar Torkamani and Volker Lohweg. Survey on time series motif discovery. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 7(2):e1199, 2017.
- Mainak Jas, Tom Dupré la Tour, Umut Simsekli, and Alexandre Gramfort. Learning the morphology of brain signals using alpha-stable convolutional sparse coding. *Advances in Neural Information Processing Systems*, 30, 2017.
- Charles Truong and Thomas Moreau. Convolutional sparse coding for time series via a penalty: An efficient algorithm with statistical guarantees. *Statistical Analysis and Data Mining: An ASA Data Science Journal*, 17(6):e70000, 2024. doi: <https://doi.org/10.1002/sam.70000>. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/sam.70000>.
- Andrew H Song, Francisco Flores, and Demba Ba. Spike sorting by convolutional dictionary learning. *arXiv preprint arXiv:1806.01979*, 2018.
- Bahareh Tolooshams, Andrew Song, Simona Temereanca, and Demba Ba. Convolutional dictionary learning based auto-encoders for natural exponential-family distributions. In *International Conference on Machine Learning*, pages 9493–9503. PMLR, 2020.
- Koray Kavukcuoglu, Pierre Sermanet, Y-Lan Boureau, Karol Gregor, Michaël Mathieu, Yann Cun, et al. Learning convolutional feature hierarchies for visual recognition. *Advances in neural information processing systems*, 23, 2010.
- Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, Jonathan Eckstein, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine learning*, 3(1):1–122, 2011.
- Thomas Moreau, Laurent Oudre, and Nicolas Vayatis. Dicod: Distributed convolutional coordinate descent for convolutional sparse coding. In *International Conference on Machine Learning*, pages 3626–3634. PMLR, 2018.
- Sébastien Bubeck et al. Convex optimization: Algorithms and complexity. *Foundations and Trends® in Machine Learning*, 8(3-4):231–357, 2015.
- Elad Hazan and Sham Kakade. Revisiting the polyak step size. *arXiv preprint arXiv:1905.00313*, 2019.
- Nicolas Loizou, Sharan Vaswani, Issam Hadj Laradji, and Simon Lacoste-Julien. Stochastic polyak step-size for sgd: An adaptive learning rate for fast convergence. In *International Conference on Artificial Intelligence and Statistics*, pages 1306–1314. PMLR, 2021.
- Mahdi Ataee, Hadi Zayyani, Massoud Babaie-Zadeh, and Christian Jutten. Parametric dictionary learning using steepest descent. In *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 1978–1981. IEEE, 2010.
- Pierre Turquais, Endrias G Asgedom, Walter Söllner, and Leiv Gelius. Parabolic dictionary learning for seismic wavefield reconstruction across the streamers. *Geophysics*, 83(4):V263–V282, 2018.
- Alekh Agarwal, Animashree Anandkumar, Prateek Jain, Praneeth Netrapalli, and Rashish Tandon. Learning sparsely used overcomplete dictionaries. In *Conference on Learning Theory*, pages 123–137. PMLR, 2014.
- Saiprasad Ravishankar, Anna Ma, and Deanna Needell. Analysis of fast structured dictionary learning. *Information and Inference: A Journal of the IMA*, 9(4):785–811, 2020.
- Sanjeev Arora, Rong Ge, and Ankur Moitra. New algorithms for learning incoherent and overcomplete dictionaries. In *Annual Conference Computational Learning Theory*, 2013.
- Lei Nie. Convergence rate of mle in generalized linear and nonlinear mixed-effects models: Theory and applications. *Journal of Statistical Planning and Inference*, 137(6):1787–1804, 2007.
- Charles Truong, Rémi Barrois-Müller, Thomas Moreau, Clément Provost, Aliénor Vienne-Jumeau, Albane Moreau, Pierre-Paul Vidal, Nicolas Vayatis, Stéphane Buffat, Alain Yelnik, Damien Ricard, and Laurent Oudre. A Data Set for the Study of Human Locomotion with Inertial Measurements Units. *Image Processing On Line*, 9:381–390, 2019.
- Niklas König, William R Taylor, Christian R Baumann, Nicole Wenderoth, and Navrag B Singh. Revealing the quality of movement: a meta-analysis review to quantify the thresholds to pathological variability during standing and walking. *Neuroscience & Biobehavioral Reviews*, 68:111–119, 2016.
- Guoqing Zheng, Yiming Yang, and Jaime Carbonell. Efficient shift-invariant dictionary learning. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 2095–2104, 2016.

Morten Mørup, Mikkel N Schmidt, and Lars K Hansen. Shift invariant sparse coding of image and music data. *Submitted to Journal of Machine Learning Research*, 5, 2008.

Paris Smaragdis. Convulsive speech bases and their application to supervised speech separation. *IEEE Transactions on Audio, Speech, and Language Processing*, 15(1):1–12, 2006.

Paul D O’Grady and Barak A Pearlmutter. Discovering speech phones using convulsive non-negative matrix factorisation with a sparseness constraint. *Neurocomputing*, 72(1-3):88–101, 2008.

Wenwu Wang, Andrzej Cichocki, and Jonathon A Chambers. A multiplicative algorithm for convulsive non-negative matrix factorization based on squared euclidean distance. *IEEE Transactions on Signal Processing*, 57(7):2858–2864, 2009.

Vikram Ramanarayanan, Louis Goldstein, and Shrikanth S Narayanan. Spatio-temporal articulatory movement primitives during speech production: Extraction, interpretation, and validation. *The Journal of the Acoustical Society of America*, 134(2):1378–1394, 2013.

Aaron Van Den Oord, Oriol Vinyals, et al. Neural discrete representation learning. *Advances in neural information processing systems*, 30, 2017.

Yong Sheng Soh. Group invariant dictionary learning. *IEEE Transactions on Signal Processing*, 69:3612–3626, 2021.

Sara Shaheen, Lama Affara, and Bernard Ghanem. Constrained convolutional sparse coding for parametric based reconstruction of line drawings. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4424–4432, 2017.

Canh T. Dinh, Nguyen H. Tran, and Tuan Dung Nguyen. Personalized federated learning with moreau envelopes. *Advances in Neural Information Processing Systems*, 33:21394–21405, 2020.

Qiushi Huang, Wenqian Jiang, Jian Shi, Chenye Wu, Dan Wang, and Zhu Han. Federated shift-invariant dictionary learning enabled distributed user profiling. *IEEE Transactions on Power Systems*, 2023.

Kaan Ozkara, Bruce Huang, and Suhas Diggavi. Personalized pca for federated heterogeneous data. In *2023 IEEE International Symposium on Information Theory (ISIT)*, pages 168–173. IEEE, 2023.

Naichen Shi and Raed Al Kontar. Personalized pca: Decoupling shared and unique features. *Journal of Machine Learning Research*, 25(41):1–82, 2024.

Jiuyun Hu, Naichen Shi, Raed Al Kontar, and Hao Yan. Personalized tucker decomposition: Modeling

commonality and peculiarity on tensor data. *arXiv preprint arXiv:2309.03439*, 2023.

Naichen Shi, Raed Al Kontar, and Salar Fattah. Heterogeneous matrix factorization: When features differ by datasets. *arXiv preprint arXiv:2305.17744*, 2023.

Geyu Liang, Naichen Shi, Raed Al Kontar, and Salar Fattah. Personalized dictionary learning for heterogeneous datasets. *Advances in Neural Information Processing Systems*, 36, 2024.

Titouan Vayer, Romain Tavenard, Laetitia Chapel, Nicolas Courty, Rémi Flamary, and Yann Soullard. Time series alignment with global invariances. *Transactions on Machine Learning Research Journal*, 2022.

Benton Jamison, Steven Orey, and William Pruitt. Convergence of weighted averages of independent random variables. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 4(1):40–44, 1965.

Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*, pages 1273–1282. PMLR, 2017.

Jingwei Zhang, Tongliang Liu, and Dacheng Tao. Going deeper, generalizing better: An information-theoretic view for deep learning. *IEEE Transactions on Neural Networks and Learning Systems*, 2023.

Patrick Wagner, Nils Strodthoff, Ralf-Dieter Boushelot, Dieter Kreiseler, Fatima I Lunze, Wojciech Samek, and Tobias Schaeffter. Ptib-xl, a large publicly available electrocardiography dataset. *Scientific data*, 7(1):1–15, 2020.

Nils Strodthoff, Patrick Wagner, Tobias Schaeffter, and Wojciech Samek. Deep learning for ecg analysis: Benchmarks and insights from ptb-xl. *IEEE journal of biomedical and health informatics*, 25(5):1519–1528, 2020.

Checklist

1. For all models and algorithms presented, check if you include:
 - (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. **Yes.** A clear description of the mathematical setting, assumptions and algorithm are presented throughout Section 2.
 - (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. **Yes.** The convergence rate of the method is presented in Section 2.5 and detailed in Appendix A. The complexity of the algorithm (the federated learning version) is given in Section 1 and detailed in Appendix B.
 - (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. **Yes.** Source code for the methods presented will be made fully available online.
 2. For any theoretical claim, check if you include:
 - (a) Statements of the full set of assumptions of all theoretical results. **Yes.** The no-overlap between atoms assumption is first presented at the end of Section 1 and then at the beginning of Section 2.5. The second assumption (known activations) is presented at the beginning of Section 2.5. The final assumption (known personalization parameters) is given in Theorem 1.
 - (b) Complete proofs of all theoretical results. **Yes.** Informal proofs are provided in the main text (Section 2.5), and complete proofs in the appendix (Appendix A).
 - (c) Clear explanations of any assumptions. **Yes.** The no-overlap between atoms assumption is discussed at the end of Section 1. The second assumption (known activations) is explained during its presentation (Section 2.5). The final assumption (known personalization parameters) is discussed below Theorem 1.
 3. For all figures and tables that present empirical results, check if you include:
 - (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). **Yes.** All details needed to reproduce the numerical experiments are given in Appendix C, with links to the datasets. Furthermore, code for the methods introduced in this work will be made available as a URL.
 - (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). **Yes.** All details regarding the numerical experiments can be found in the main text (Section 3) and in Appendix C.
 - (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). **Yes.** Information on the measure and error bars is given alongside each figure.
 - (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). **Yes.** Information on the computing infrastructure used is given in Appendix C.
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:
 - (a) Citations of the creator If your work uses existing assets. **Yes.** The authors of the CSC optimization scheme (Truong and Moreau, 2024) and the gait segmentation method (Voisard et al., 2024) were cited.
 - (b) The license information of the assets, if applicable. **Not Applicable.**
 - (c) New assets either in the supplemental material or as a URL, if applicable. **Yes.** The authors code for the methods discussed in the paper will be available via a URL.
 - (d) Information about consent from data providers/curators. **Not Applicable.**
 - (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. **Not Applicable.**
 5. If you used crowdsourcing or conducted research with human subjects, check if you include:
 - (a) The full text of instructions given to participants and screenshots. **Not Applicable.**
 - (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. **Not Applicable.**
 - (c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. **Not Applicable.**

A COMPLETE PRESENTATION OF THEORETICAL GUARANTEES

A.1 Rotations

Section 2.3 mentioned a complexity of the IPU step in $\mathcal{O}(SKMNPT_{\text{grad}})$ for the general case, with T_{grad} the number of gradient steps. However, this complexity can be decreased if more structure is assumed on f .

Let us consider the particular case of orthogonal transformations. Some multivariate time series, for example hand-writing-related signals (Vayer et al., 2022), naturally exhibit rotations in space. To model such data in the PerCDL framework, let us define the rotation transformation of ϕ , $f_r(\phi, \mathbf{a}) = \mathbf{a}\phi$, for any $\phi, \mathbf{a} \in (\mathbb{R}^P)^L \times \mathcal{O}_P$. Let \mathbf{x}^s be a signal whose segments with non-zero activation are written X_i . Lemma 2 shows that the IPU update can be solved directly.

Lemma 2. *Given an atom $\phi \in (\mathbb{R}^P)$ and raw data $X_i \in (\mathbb{R}^P)$ for any $i \in [Q]$, where Q is the number of repeated noisy patterns in an individual time series, the solution of $\underset{\mathcal{O} \in \mathcal{O}_P}{\operatorname{argmin}} \sum_{i=1}^m |X_i - O\phi|^2$ is VU^\top where $\phi \sum_{i=1}^m X_i^\top / m = U\Sigma V^\top$ is a singular values decomposition.*

Proof. We have,

$$\sum_{i=1}^m |X_i - O\phi|^2 = \sum_{i=1}^m |X_i|^2 + |O\phi|^2 - 2 \operatorname{tr}(X_i^\top O\phi) \quad (8)$$

$$= \left[\sum_{i=1}^m |X_i|^2 + |\phi|^2 \right] - 2 \operatorname{tr} \left(\phi \sum_{i=1}^m X_i^\top O \right), \quad (9)$$

where we used that $|O\phi|^2 = |\phi|^2$ since $O \in \mathcal{O}_P$. Then,

$$\underset{\mathcal{O} \in \mathcal{O}_P}{\operatorname{argmax}} \operatorname{tr} \left(\phi \sum_{i=1}^m X_i^\top O \right) = \underset{\mathcal{O} \in \mathcal{O}_P}{\operatorname{argmin}} \sum_{i=1}^m |X_i - O\phi|^2, \quad (10)$$

and

$$\operatorname{tr} \left(\phi \sum_{i=1}^m X_i^\top O \right) = \operatorname{tr} (\Sigma V^\top OU), \quad (11)$$

which gives $O = VU^\top$ since Σ is a non-negative diagonal and O is orthogonal. \square

Given Lemma 2, the complexity of IPU drops from $\mathcal{O}(SKMNPT_{\text{grad}})$ to $\mathcal{O}(SKP^3)$. A similar update rule was shown in Soh (2021) for group-invariant DL.

A.2 Toward Mixed-Effects Models

A.2.1 Alternative Formulation of Lemma 1 and Proof

Section 2.5 introduced **H1**, the foundation hypothesis used to derive a convergence rate for the MLE on the common atom ϕ_* . Lemma 1 then proved that **H1** is met most of the time in practice.

In what follows, we propose an alternative, more formal statement of this Lemma.

Lemma 3. *Assumption H1 is met in the two following cases:*

- (*Time warping transformations*) For any $\mathbf{a} \in \Theta \subset \mathbb{R}^M$, $\mathcal{L}(\mathbf{a}) = \mathcal{T}_{\psi_{\mathbf{a}}, \sigma}$ where $\mathcal{T}_{\cdot, \sigma}$ and ψ are defined in Equations (4) and (5) with σ small enough. For any $s \in [S]$, $\mathbf{a}_*^s \stackrel{i.i.d.}{\sim} \mathcal{P}$ where \mathcal{P} is a distribution supported on Θ such that for any $l \in [L]$,

$$S_l = \sum_{i=1}^L \mathbb{P}(L\psi(a)(i/L) \in (l - 1/2, l + 1/2)) > 0. \quad (12)$$

- (*Rotations*) For any $\mathbf{a} \in \Theta \subset \mathbb{R}^M$, $\mathcal{L}(\mathbf{a}) \in \mathcal{O}_P(\mathbb{R})$.

Proof. The case of rotations is trivial since for any $s \in [S]$, $\mathcal{L}(\mathbf{a}_*^s)^\top \mathcal{L}(\mathbf{a}_*^s) = I_L$.

The case of discrete temporal reparameterizations is less obvious. For any $\mathbf{x} \in (\mathbb{R}^P)^L$, we have,

$$(\mathbb{E}_{\mathbf{a} \sim \mathcal{P}} (\mathcal{T}_{\psi_{\mathbf{a}}, \sigma}^\top \mathcal{T}_{\psi_{\mathbf{a}}, \sigma}) \mathbf{x})^\top \mathbf{x} = \mathbb{E} ((\mathcal{T}_{\psi_{\mathbf{a}}, \sigma}^\top \mathcal{T}_{\psi_{\mathbf{a}}, \sigma} \mathbf{x})^\top \mathbf{x}) = \mathbb{E} ((\mathcal{T}_{\psi_{\mathbf{a}}, \sigma} \mathbf{x})^\top \mathcal{T}_{\psi_{\mathbf{a}}, \sigma} \mathbf{x}). \quad (13)$$

We have $\mathbb{E} ((\mathcal{T}_{\psi_{\mathbf{a}}, \sigma} \mathbf{x})^\top \mathcal{T}_{\psi_{\mathbf{a}}, \sigma} \mathbf{x}) = \sum_{i=1}^L \mathbb{E} ([\mathcal{T}_{\psi_{\mathbf{a}}, \sigma} \mathbf{x}]_i^2)$. For any $i \in [L]$, denoting by $U_l = x_l + (L\psi_{\mathbf{a}}(i/L) - l)(x_{l+1} - x_l)$,

$$\mathbb{E} ([\mathcal{T}_{\psi_{\mathbf{a}}, \sigma} \mathbf{x}]_i^2) = \mathbb{E} \left(\left(\sum_{l=1}^L w_l(\psi_{\mathbf{a}}, \sigma, i) U_l \right)^2 \right) \quad (14)$$

$$= \mathbb{E} \left(\sum_{l=1}^L \sum_{m=1}^L w_l(\psi_{\mathbf{a}}, \sigma, i) w_m(\psi_{\mathbf{a}}, \sigma, i) U_l U_m \right) \quad (15)$$

$$= \sum_{l=1}^L \sum_{m=1}^L \underbrace{\mathbb{E} (w_l(\psi_{\mathbf{a}}, \sigma, i) w_m(\psi_{\mathbf{a}}, \sigma, i) U_l U_m)}_{=\Delta_{k,m,i}}. \quad (16)$$

With $C_l = (l - \frac{1}{2}, l + \frac{1}{2})$ for any $l \in [L]$, note that $w_l(\psi_{\mathbf{a}}, \sigma, i) \approx \mathbb{1}_{C_l}(L\psi_{\mathbf{a}}(i/L))$ for any $i \in [L]$. Therefore, computing $\Delta_{l,m,i}^* = \mathbb{E} (\mathbb{1}_{C_l}(L\psi_{\mathbf{a}}(i/L)) \mathbb{1}_{C_m}(L\psi_{\mathbf{a}}(i/L)) U_l U_m)$ for any $(l, m, i) \in [L]^3$ is interesting. The proof of $\Delta_{l,m,i}^* \approx \Delta_{l,m,i}$ is postponed to the end of the current proof.

We have,

$$\Delta_{l,m,i}^* = \mathbb{E} (\mathbb{1}_l(m) \mathbb{1}_{C_l}(L\psi_{\mathbf{a}}(i/L)) U_l^2) \quad (17)$$

$$= \mathbb{1}_l(m) \mathbb{E} (U_l^2 | L\psi_{\mathbf{a}}(i/L) \in C_l) \mathbb{P}(L\psi_{\mathbf{a}}(i/L) \in C_l). \quad (18)$$

Then, denoting by $P_{i,l} = \mathbb{P}(L\psi_{\mathbf{a}}(i/L) \in C_l)$, $S_l = \sum_{i=1}^L P_{i,l} > 0$ (by (12)) and $\bar{P}_{i,l} = P_{i,l}/S_l$ for any $(i, l) \in [L]^2$,

$$\sum_{i=1}^L \sum_{l=1}^L \sum_{m=1}^L \Delta_{l,m,i}^* = \sum_{i=1}^L \sum_{l=1}^L \mathbb{E} (U_l^2 | L\psi_{\mathbf{a}}(i/L) \in C_l) P_{i,l} \quad (19)$$

$$\geq \sum_{l=1}^L S_l \sum_{i=1}^L [x_l(1 - \alpha_{i,l}) + x_{l+1}\alpha_{i,l}]^2 \frac{P_{i,l}}{S_l}, \quad (20)$$

where we applied the Jensen inequality and wrote $\alpha_{i,l} = \mathbb{E}(L\psi_{\mathbf{a}}(i/L) - l | L\psi_{\mathbf{a}}(i/L) \in C_l)$ for any $(i, l) \in [L]^2$. Note that $\alpha_{i,l} \in (-1/2, 1/2)$ since the expectation is taken on the event $L\psi_{\mathbf{a}}(i/L) \in C_l$.

$$\sum_{i=1}^L \sum_{l=1}^L \sum_{m=1}^L \Delta_{l,m,i}^* \geq \sum_{l=1}^L S_l \sum_{i=1}^L (x_l(1 - \alpha_{i,l}) + x_{l+1}\alpha_{i,l})^2 \bar{P}_{i,l} \quad (21)$$

$$\geq \sum_{l=1}^L S_l \left(x_l(1 - \sum_{i=1}^L \bar{P}_{i,l}\alpha_{i,l}) + x_{l+1} \sum_{i=1}^L \bar{P}_{i,l}\alpha_{i,l} \right)^2 = \mathbf{x}^\top \bar{D}^\top \mathbf{S} \bar{D} \mathbf{x}, \quad (22)$$

where

$$B_1 = \begin{pmatrix} B_1 & \sum_{i=1}^L \bar{P}_{i,1}\alpha_{i,1} \\ 0 & \left(1 - \sum_{i=1}^L \bar{P}_{i,2}\alpha_{i,2}\right) \\ 0 & 0 \end{pmatrix}, \quad \bar{D} = \begin{pmatrix} 0 & (0) \\ \sum_{i=1}^L \bar{P}_{i,2}\alpha_{i,2} & (0) \\ \dots & \dots \end{pmatrix} \quad (23)$$

$$\mathbf{S} = \text{diag}((S_l)_{l \in [L]}). \quad (24)$$

\mathbf{S} is a positive diagonal matrix by (12) and \bar{D} is a bi-diagonal matrix with a positive diagonal, thus $\mathbf{x}^\top \bar{D}^\top \mathbf{S} \bar{D} \mathbf{x} > 0$ is positive.

Then, coming back to $\Delta_{l,m,i}^* \approx \Delta_{l,m,i}$, for any $x \in (\mathbb{R}^P)^L$,

$$\sum_{i=1}^L \sum_{l=1}^L \sum_{m=1}^L \Delta_{k,m,i} = x^\top \tilde{W}_\sigma x \quad (25)$$

where \tilde{W}_σ is a matrix with coordinates which are functions of element of the form $\mathbb{E}(w_l(\psi_a, \sigma, i) w_m(\psi_a, \sigma, i) E_{i,l,m})$ with $|E_{i,l,m}| \leq L^2$. Using that $w_l(\psi_a, \sigma, i) \in [0, 1]$ converges almost surely toward $\mathbb{1}_{C_l}(L\psi_a(i/L))$ as $\sigma \rightarrow 0^+$, we have, using the Dominated Convergence Theorem, that \tilde{W}_σ converges toward a matrix \tilde{W}_* such that:

$$x^\top \tilde{W}_* x = \sum_{i=1}^L \sum_{l=1}^L \sum_{m=1}^L \Delta_{l,m,i}^* \geq x^\top \bar{D}^\top \mathbf{S} \bar{D} x > 0 \quad (26)$$

Therefore, since the space of invertible matrix is open, there exists $\sigma_0 > 0$ such that for any $\sigma > \sigma_0$, we have that $\mathbb{E}_{A \sim \mathcal{P}}(\mathcal{T}_{\psi_a, \sigma}^\top \mathcal{T}_{\psi_a, \sigma})$ is a symmetric positive definite matrix. \square

A.2.2 Circling Back to Lemma 1

Let us define $\mathsf{T}_\psi^l = (\exists i \in [L] : L\psi_a(i/L) \in (l - 1/2, l + 1/2))$ for any $l \in [L]$. For any $\epsilon > 0$,

$$P_{Id}^\epsilon = \mathbb{P}(\cap_{i \in [L]} (L\psi_a(i/L) \in (i - \epsilon, i + \epsilon))) \leq \mathbb{P}(\cap_{l=1}^L \mathsf{T}_\psi^l)^L \leq \prod_{l=1}^L \mathbb{P}(\mathsf{T}_\psi^l) \leq \prod_{l=1}^L S_l, \quad (27)$$

and $P_{Id}^\epsilon > 0$ as long as \mathcal{P} has enough mass around zero since $(\mathbf{a} = 0) = (\psi_a = \text{Id})$.

Therefore, condition (12) from Lemma 3 can be considered as very weak. Informally, on the event $(L\psi_a(i/L) \in (l - 1/2, l + 1/2))$, we have $[\mathcal{T}_{\psi_a, \sigma}(\phi)]_i \approx \phi_l + (L\psi_a(i/L) - l)(\phi_{l+1} - \phi_l) \approx \phi_l$ as $\sigma \ll 1$.

Thus, equation (12) states that, for any $l \in [L]$, the l -th sample of the common atom ϕ has a positive probability to be seen in a personalized atom ϕ_*^s . This statement is the one presented in Lemma 1 from the main text.

A.2.3 Proof of Theorem 1

For completeness sake, this section will reason on a slightly more complete version of Theorem 1. The following exposition naturally directly transposes to Theorem 1.

Theorem 2. Under **H1**, **H2**, and assuming that the true parameters $(\mathbf{a}_s^*)_{s \in [S]}$ are given, there exists almost surely an integer \tilde{S} , such that for any $S \geq \tilde{S}$, $\Sigma_{S,p} = \sum_{s=1}^S p_s \mathcal{L}(\mathbf{a}_s^*)^\top \mathcal{L}(\mathbf{a}_s^*)$ is invertible. Then, for any $S \geq \tilde{S}$, the MLE of the common atom ϕ_* is

$$\hat{\phi}_{mle} = (\Sigma_{S,p})^{-1} Y_\Sigma, \quad Y_\Sigma = \sum_{s=1}^S \sum_{j=1}^{p_s} \mathcal{L}(\mathbf{a}_s^*)^\top [j] y^s[j], \quad (28)$$

and thus we have $\hat{\phi}_{mle} \sim \mathcal{N}(\phi_*, (\Sigma_{S,p})^{-1})$. Moreover, $\hat{\phi}_{mle}$ converges toward ϕ_* at a rate of $\mathcal{O}\left(1/\rho\sqrt{\sum_{s=1}^S p_s}\right)$ in probability.

Proof. First, by (Jamison et al., 1965, Theorem 3), which is an extension of the Strong Law of Large Numbers (SLLN) to the weighted averages, we have that almost surely $\Sigma_{S,p} / \sum_{s=1}^S p_s \rightarrow \mathbb{E}_{\mathcal{P}}(\mathcal{L}(\mathbf{a}_*)^\top \mathcal{L}(\mathbf{a}_*))$ as S goes to infinity. Indeed, (Jamison et al., 1965, Theorem 3) states that SLLN applies if and only if $\limsup_{x \rightarrow \infty} \mathsf{S}(x)/x < \infty$ with $\mathsf{S}(x) = \#\{n \geq 0 : \sum_{s=1}^n p_s/p_n \leq x\}$, $(\mathbf{a}_s^*)_{s \in [S]}$ i.i.d. and $\mathbb{E}_{\mathcal{P}}(|\mathcal{L}(\mathbf{a}_*)^\top \mathcal{L}(\mathbf{a}_*)|) < \infty$. This condition is met since for any $n \in [S]$, $\mathsf{S}(\sum_{s=1}^n p_s/(C+1)) \leq n$ because by **H2**, for any $n \in [N]$,

$$\sum_{s=1}^n p_s/C \leq \sum_{s=1}^n p_s/p_n \leq \sum_{s=1}^n p_s. \quad (29)$$

Thus, for any $n \in [S]$, $(C+1)\mathsf{S}(\sum_{s=1}^n p_s/(C+1))/\sum_{s=1}^n p_s \leq C+1$. This ensures $\limsup_{x \rightarrow \infty} \mathsf{S}(x)/x < \infty$ since $x \rightarrow \mathsf{S}(x)$ is not decreasing.

Moreover, the set of invertible matrices of $\mathcal{M}_{L,L}(\mathbb{R})$ is open. Therefore, there exists almost surely \tilde{S} such that, for any $S \geq \tilde{S}$, $\Sigma_{S,p}/\sum_{s=1}^S p_s$ is invertible since $\mathbb{E}_P \left(\mathcal{L}(\mathbf{a}_*^s)^\top \mathcal{L}(\mathbf{a}_*^s) \right)$ is invertible (**H1**). In addition, $\Sigma_{S,p}$ is symmetric positive as a sum of symmetric positive matrices. Therefore, for any $S \geq \tilde{S}$, $\Sigma_{S,p}$ is symmetric definite positive and the minimum eigen values of $\Sigma_{S,p}/\sum_{s=1}^S p_s$ converges to ρ almost surely.

Secondly, the log likelihood related to ϕ_* in (7) is

$$F : \phi_* \in (\mathbb{R}^P)^L \mapsto \sum_{s=1}^S \sum_{j=1}^{p_s} |y^s[j] - \mathcal{L}(\mathbf{a}_*^s) \phi_*|^2 / (2\sigma^2) + L \log(2\pi\sigma^2) / 2. \quad (30)$$

Then, writing $\nabla F(\hat{\Phi}) = 0$ yields,

$$\Sigma_{S,p} \hat{\phi}_{mle} = Y_\Sigma. \quad (31)$$

For any $S \geq \tilde{S}$, $\Sigma_{S,p}$ is invertible. Therefore, we have (28).

Hence,

$$\hat{\phi}_{mle} - \phi_* = (\Sigma_{S,p})^{-1} Y_\Sigma - \phi_* \quad (32)$$

$$= (\Sigma_{S,p})^{-1} \left[\sum_{s=1}^S p_s \mathcal{L}(\mathbf{a}_*^s)^\top \mathcal{L}(\mathbf{a}_*^s) \phi_* \right] - \phi_* + (\Sigma_{S,p})^{-1} \sum_{s=1}^S \sum_{j=1}^{p_s} \mathcal{L}(\mathbf{a}_*^s)^\top [j] \epsilon^s[j] \quad (33)$$

$$= \sum_{s=1}^S \sum_{j=1}^{p_s} (\Sigma_{S,p})^{-1} \mathcal{L}(\mathbf{a}_*^s)^\top [j] \epsilon^s[j]. \quad (34)$$

We sum independent normal variables of covariance matrix $\sigma^2 (\Sigma_{S,p})^{-1} \mathcal{L}(\mathbf{a}_*^s)^\top \left[(\Sigma_{S,p})^{-1} \mathcal{L}(\mathbf{a}_*^s) \right]^\top$, which yields,

$$\hat{\phi}_{mle} - \phi_* \sim \mathcal{N} \left(0, (\Sigma_{S,p})^{-1} \right). \quad (35)$$

Asymptotically, the largest eigen value of $(\Sigma_{S,p})^{-1}$ is $1/\left(\rho \sum_{s=1}^S p_s\right)$, resulting in a convergence rate of $1/\left(\rho \sum_{s=1}^S p_s\right)$ in probability. \square

B FEDERATED LEARNING META-ALGORITHM

The expression of the MLE from Theorem 2 was used to adapt the meta-algorithm algorithm 1 to a Federated Learning framework under the non-overlapping assumption. In Algorithm 2, the Global Server aggregates information from the Local Servers through a weighted average. Each communication, from the Local Servers to the Central Server or vice versa, has a memory footprint of $\mathcal{O}(L)$ since the transmitted vectors have nearly the same shape as the common atom. At each step, the complexity on the Local Server is $\mathcal{C}(\text{IPU}) + \mathcal{C}(\text{CSC}) = \mathcal{O}(\max(KMT_{grad}N, KLN(\log(N) + P))$. In our experiments, $M = DW \approx 50 < L = 100$ and $P < 20$. Moreover, $n_{init} = n_{perso} = 5$ is enough to reach convergence. Hence a final complexity per Local Server of $\mathcal{O}(N \log(N))$.

Robustness improvement The personalization step might not be robust when the individual parameters are poorly estimated (recall that the update formula of the personalized common atom in (28) assumes a perfect estimation of the individual parameters). To increase its robustness, we follow (McMahan et al., 2017) to replace the update on the Local Servers by a gradient descent on both the individual parameters \mathbf{a}^s and the personalized dictionaries $\tilde{\phi}_k^s$ at the same time. The common dictionary can then be updated as $\phi_k \leftarrow \sum_{s=1}^S \tilde{\phi}_k^s / S$ for any $k \in [K]$.

In order to avoid over-adaptation of the personalized dictionaries, the individual parameters should be initialized with a preliminary step of IPU before running the joint update on each Local Server. Such amendments result

```

1 Inputs  $\mathbf{x}^s \in \mathbb{R}^{S \times N \times P}$ ,  $f : (\mathbb{R}^P)^L \times \mathbb{R}^M \mapsto (\mathbb{R}^P)^L$ 
2 Outputs  $\Phi \in \mathbb{R}^{K \times L \times P}$ ,  $\mathbf{A} \in \mathbb{R}^{S \times K \times M}$ ,  $\mathbf{Z} \in \mathbb{R}^{S \times K \times N-L+1}$ 
1  $\Phi, \mathbf{A}, \mathbf{Z} \leftarrow \text{setInitialValues}()$ , for any  $k \in [K]$ ,  $c_k = 1$  // Initialization
// First estimates
2 while  $i \leq n_{init}$  do
3   In each Local Server for  $1 \leq s \leq S$  do
4     | For any  $k \in [K]$ ,  $\mathbf{z}_k^s \leftarrow \mathbf{z}_k^s \times c_k$  ;
5     |  $\mathbf{z}^s \leftarrow \text{CSC}(\mathbf{x}^s, \mathbf{z}^s, \Phi)$ ;
6     | For any  $k \in [K]$ ,  $\tilde{\phi}_k^s \leftarrow \sum_{j=1}^{p_s^k} y_k^s[j] / p_s^k$  (Individual Barycenter) and  $\tilde{\phi}_k^s$  is sent to the global
      | server
7   end
8   [Central Server];
9   For any  $k \in [K]$ ,  $\phi_k \leftarrow \sum_{s=1}^S \tilde{\phi}_k^s / S$  (Global Barycenter),  $\phi_k, c_k \leftarrow \phi_k / |\phi_k|, |\phi_k|$  (Normalization),
   | and  $\Phi = (\phi_k)_{k \in [K]}, (c_k)_k$  are sent to each local server.
10 end
// Personalization
11 while  $i \leq n_{perso}$  do
12   In each Local Server for  $1 \leq s \leq S$  do
13     |  $\mathbf{z}^s \leftarrow \text{CSC}(\mathbf{x}^s, \mathbf{z}^s, \Phi)$ ;
14     |  $\mathbf{A}_s \leftarrow \text{IPU}(\mathbf{X}_s, \mathbf{z}^s, \Phi, \mathbf{A}_s, f)$ ; For any  $k \in [K]$ ,  $\tilde{\Sigma}_s^k \leftarrow p_s^k \mathcal{L}(\mathbf{a}_k^s)^\top \mathcal{L}(\mathbf{a}_k^s)$ ,
      |  $\tilde{\phi}_k^s \leftarrow \sum_{j=1}^{p_s^k} \mathcal{L}(\mathbf{a}_k^s)^\top y_k^s[j]$  and  $p_s^k, \mathbf{a}_k^s, \tilde{\phi}_k^s$  is sent to the global server
15   end
16   [Central Server]
17   For any  $k \in [K]$ ,  $\tilde{\Sigma}_k \leftarrow \sum_{s=1}^S p_s^k \mathcal{L}(\mathbf{a}_k^s)^\top \mathcal{L}(\mathbf{a}_k^s)$ ,  $\phi_k \leftarrow \tilde{\Sigma}_k^{-1} \sum_{s=1}^S \tilde{\phi}_k^s$  (Weighted Global
      | Barycenter),
18   |  $\phi_k, c_k \leftarrow \phi_k / |\phi_k|, |\phi_k|$  (Normalization), and  $\Phi = (\phi_k)_{k \in [K]}, (c_k)_k$  are sent to each local server.
19 end

```

Algorithm 2: A Federated Learning algorithm to find an approximated solution of (3). Remark that $\hat{\Phi}$ is the personalized dictionary defined in (2) and Φ is the common dictionary. We denote by $y_k^s[j]$ the j^{th} subpart of \mathbf{x}^s where a pattern is recognized by a non-zero activation in \mathbf{z}_k^s and p_s^k the number of non-zero activation in \mathbf{z}_k^s . This algorithm works under the assumptions of Section 2.5. The notations of **H1** are used.

in a final complexity per local server asymptotically identical as the one of Algorithm 2, with an even smaller communication cost.

C ADDITIONAL NUMERICAL EXPERIMENTS

Unless otherwise stated, all experiments were conducted on a computing server equipped with an Intel® Xeon® Gold 5220R processor at 2.20 GHz . The code used for the experiments can be found at the following online repository².

As is customary in the field of time series representation (Dupré la Tour et al., 2018; Power et al., 2023; Liang et al., 2024; Jas et al., 2017; Moreau et al., 2018; Mørup et al., 2008; Shaheen et al., 2017; Song et al., 2018), no data separation into training/validation/test sets was implemented. The ambition was to reconstruct a given dataset as truthfully as possible rather than investigate the generalization capabilities of the method (within-signal generalization is somewhat assumed implicitly due to the structured nature of the physiological activity recorded).

Interesting Implementations Detail. We have observed that slightly centering the common atoms after the IPU step by considering $\phi_{new} = \mathcal{T}_{\bar{\psi}, \sigma} \phi_{old}$ with $\bar{\psi} = \sum_{i=1}^S \psi_{a_i}$ has a regularizing effect: it smooths the common atom. This optional additional step was employed in the experiments.

Time Complexity. In order to give a better idea of the temporal complexity of the different methods described in this work, the next paragraph details the actual computation times observed on a personal laptops (MacBook Pro, M3 Max chip) for a subset of the experiments presented in this section.

1. Synthetic dataset ($K = 2$, $N = 500$, $n_{steps} = 5$):
 - (a) $S = 2$: IndCDL: 0.29s; PopCDL: 0.17s; PerCDL: 5.86s.
 - (b) $S = 16$: IndCDL: 2.20s; PopCDL: 0.32s; PerCDL: 7.33s.
 - (c) $S = 128$: IndCDL: 17.57s; PopCDL: 2.79s; PerCDL: 21.50s.
 - (d) $S = 1024$: IndCDL: 138.57s; PopCDL: 18.46s; PerCDL: 127.25s.
2. Gait dataset ($S = 150$, $N = 15457$, $n_{steps} = 5$): IndCDL: 26.86s; PopCDL: 5.94s; PerCDL: 48.37s.
3. ECG dataset ($S = 4465$, $N = 1000$, $n_{steps} = 20$): IndCDL: 2465.09s; PopCDL: 292.67s; PerCDL: 1731.81s.

C.1 Validation on Synthetic Data

An extensive list of experiments were conducted on synthetic data to verify the theoretical findings presented above. The toy dataset consisted of S time series of equal-length N containing $r = 3$ repetitions of $K = 2$ common pattern of length L , distributed randomly within each time series under the no overlap condition. The shared structure was transformed using random personalization parameters drawn uniformly random from $(-1, 1)^M$ and projected onto Θ in each time series with the transformation function described in section 2.4. In the remainder of this section, the personalization hyper-parameters were fixed with $D = 3$ and $W = 10$. The sparsity balance parameter was fixed at $\lambda = 10^{-2}$.

Finding the Activations. In this section, the CSC step of PerCDL was evaluated. PerCDL was initialized with the true common atoms and personalization parameters, which remained fixed for the duration of the experiment. Figure 6 presents the results of the optimization process.

Our findings suggest that PerCDL can effectively recover a set of known activations when the common atoms and the personalization parameters are fixed.

Finding the Common Atom. In this section, the CDU step of PerCDL was evaluated. PerCDL was initialized with the true activations and personalization parameters, which remained fixed for the duration of the experiment. The common dictionary was initialized with the personalized atoms of the first signal. Figure 7 presents the results of the optimization process.

Our findings suggest that PerCDL can effectively recover a set of known common atoms when the activations and the personalization parameters are fixed.

²<https://github.com/axelroques/PerCDL>.

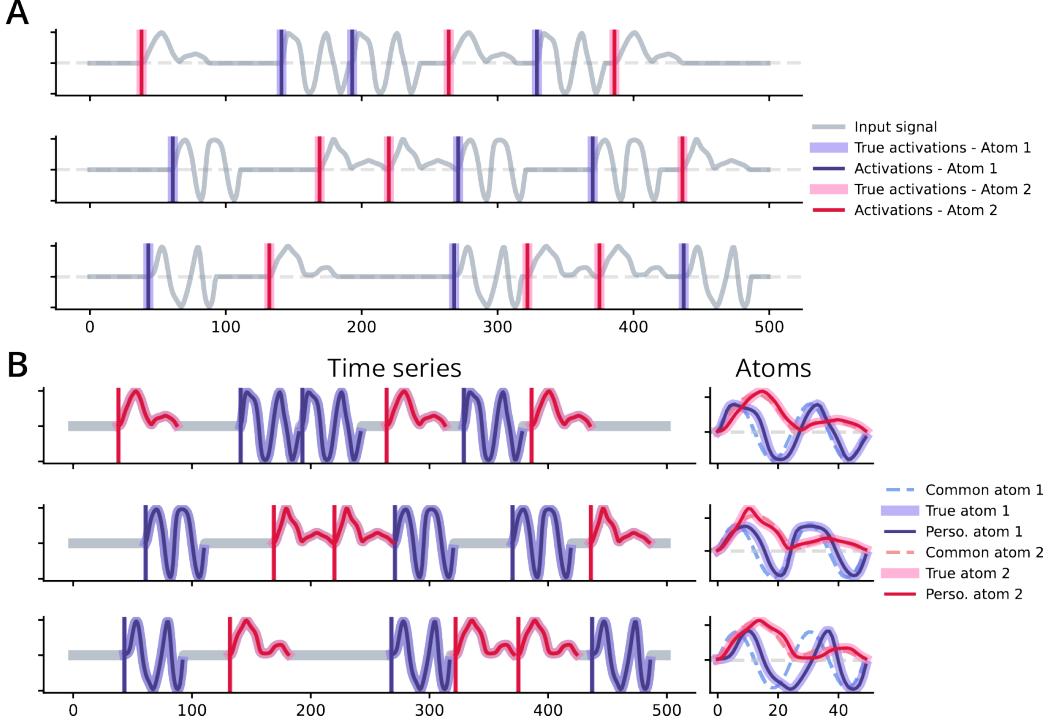


Figure 6: Recovery of the Activations with Fixed Common Atoms and Personalization Parameters.
Starting from known common atoms and personalizations parameters, PerCDL successfully recovers the true activations (A, solid and transparent vertical lines). Reconstruction of the dataset, as well as the common and personalized dictionaries are shown in B. Dataset parameters: $N = 500$, $S = 3$, $L = 50$, $D = 3$, $W = 10$. Optimization parameters: $n_{steps} = 250$, step size of 10^{-3} .

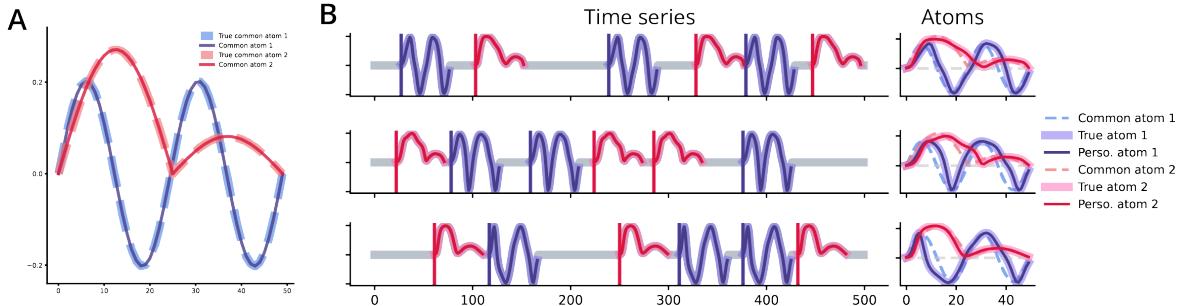


Figure 7: Recovery of the Common Atoms with Fixed Activations and Personalization Parameters.
Starting from known activations and personalization parameters, PerCDL successfully recovers the true common atoms (A, solid and dashed colored lines). Reconstruction of the dataset, as well as the common and personalized dictionaries are shown in B. Dataset parameters: $N = 500$, $S = 3$, $L = 50$, $D = 3$, $W = 10$. Optimization parameters: $n_{steps} = 250$, step size of 10^{-3} .

Finding the Personalization Parameters. In this section, the IPU step of PerCDL was evaluated. PerCDL was initialized with the true activations and common atoms, which remained fixed for the duration of the experiment. Figure 8 presents the results of the optimization process.

Our findings suggest that PerCDL can effectively recover a set of known personalization parameters when the activations and the common atoms are fixed.

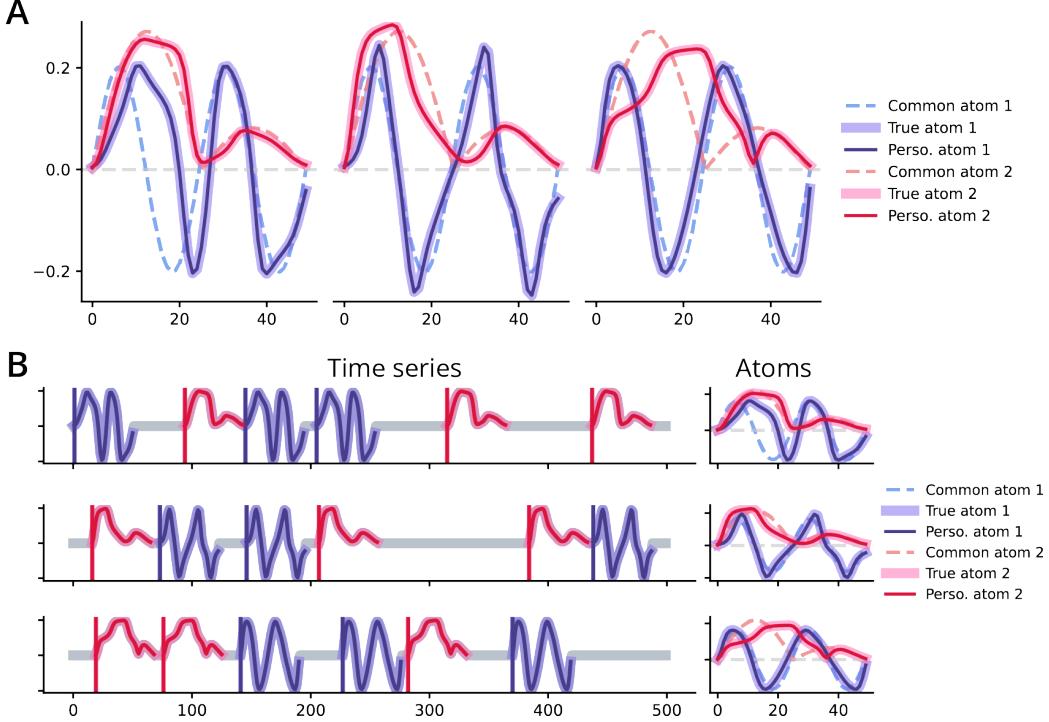


Figure 8: Recovery of the personalization Parameters with Fixed Activations and Common Atoms.
Starting from known activations and common atoms, PerCDL successfully recovers the true personalization parameters (A, solid and dashed colored lines). Reconstruction of the dataset, as well as the common and personalized dictionaries are shown in B. Dataset parameters: $N = 500$, $S = 3$, $L = 50$, $D = 3$, $W = 10$. Optimization parameters: $n_{steps} = 250$, step size of 10^{-3} .

C.2 Sensitivity Analysis

The sensitivity of the transformation f (4) with respect to its hyper-parameters D, W was investigated. A synthetic dataset was generated ($N = 500$, $S = 3$) with $(D, W) = (3, 10)$ by applying a transformation $f(\cdot, \mathbf{a})$ on the two fixed common atoms shown in figure 7, with random parameters \mathbf{a} uniformly sampled in the interval $[-1, 1]^M$. PerCDL was then initialized with the known common atoms and fixed activations, and the effect of both hyper-parameters on the reconstruction error was studied using a grid search. Results are presented in Figure 9.

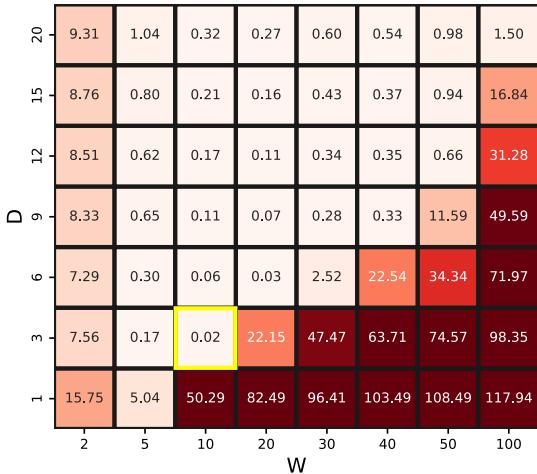


Figure 9: Sensitivity Analysis. Reconstruction error (Euclidean distance) versus the choice of time warping hyper-parameters D, W on a synthetic dataset. The correct hyper-parameters used to generate the data are $(D, W) = (3, 10)$, marked with a yellow box.

Our results indicate that larger D reduces the sensitivity to W , and generally leads to good performance. This

is unsurprising given the known benefits of over-parameterization Zhang et al. (2023).

C.3 Convergence Toward the Common Atoms

The ability of the different methods in identifying the common structures was evaluated on synthetic datasets. All datasets tested contained a varying number of time series, from $S = 1$ to $S = 1024$, of equal length $N = 1000$, with $r = 3$ repetitions of $K = 2$ common atoms of length $L = 50$, without overlap. Quantification results were presented in the main text. Figure 10 illustrates qualitatively the convergence toward the common atoms with an increase in the number of signals.

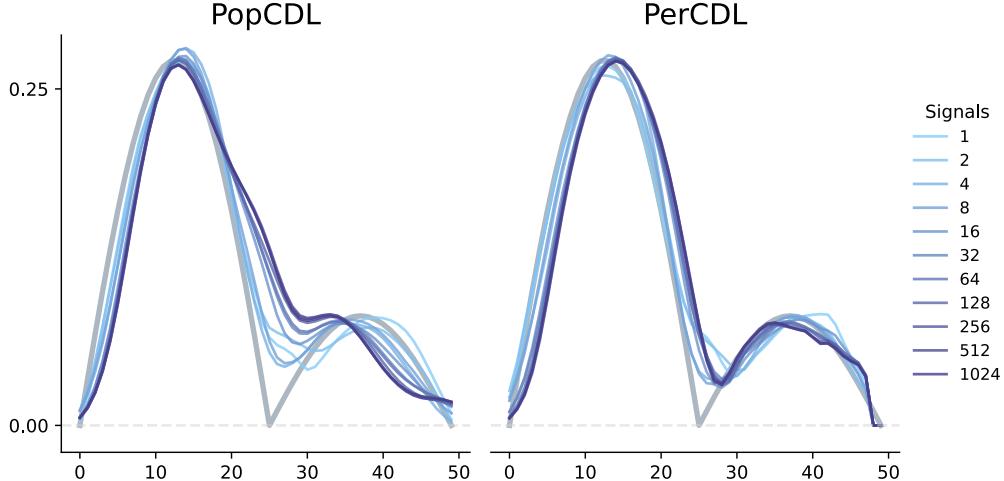


Figure 10: **Recovery of the Common Atoms in Synthetic Data as a Function of S .**

PerCDL efficiently recovers the shared information with an increase in the dataset’s size, contrary to PopCDL.

C.4 Robustness to Noise

The robustness of the different methods was evaluated on synthetic data contaminated with Gaussian or impulse noise. All synthetic datasets tested contained $S = 32$ times series of equal length $N = 1000$, with $r = 3$ repetitions of $K = 2$ common atoms of length $L = 50$, without overlap. Results with impulse noise were presented in the main text. The following paragraph will investigate the impact of Gaussian noise.

Figure 11(A) presents the evolution of the distance to the personalized atoms as a function of the signal to Gaussian noise ratio (SNR). Two distance metrics were tested: the Euclidean distance and the DTW. Only the latter was included to ensure a fair comparison between methods despite potential discrepancies in temporal offset between the true and the learned atoms, which may arise due to different activation matrices. We observe that no matter the metric or the SNR level, the atoms learned through PopCDL and PerCDL are closer to the true underlying atoms compared to IndCDL under noisy conditions ($SNR < 20$).

Figure 11(B) presents the evolution of the reconstruction error—computed as the Euclidean distance between the input signal and the reconstruction using the activations and atoms identified following the optimization process—as a function of the SNR. Our results indicate that both PopCDL and PerCDL degrade relatively less than IndCDL under high noise levels ($SNR < 20$).

C.5 Application to Locomotion Data

The interested reader may find the complete dataset used for the experiment at the following online repository³ and a detailed description of the experiment conducted in the original article (Truong et al., 2019). Briefly, three groups of participants (Healthy, Orthopedic, and Neurological) underwent the same basic walking protocol

³<https://github.com/deepcharles/gait-data>.

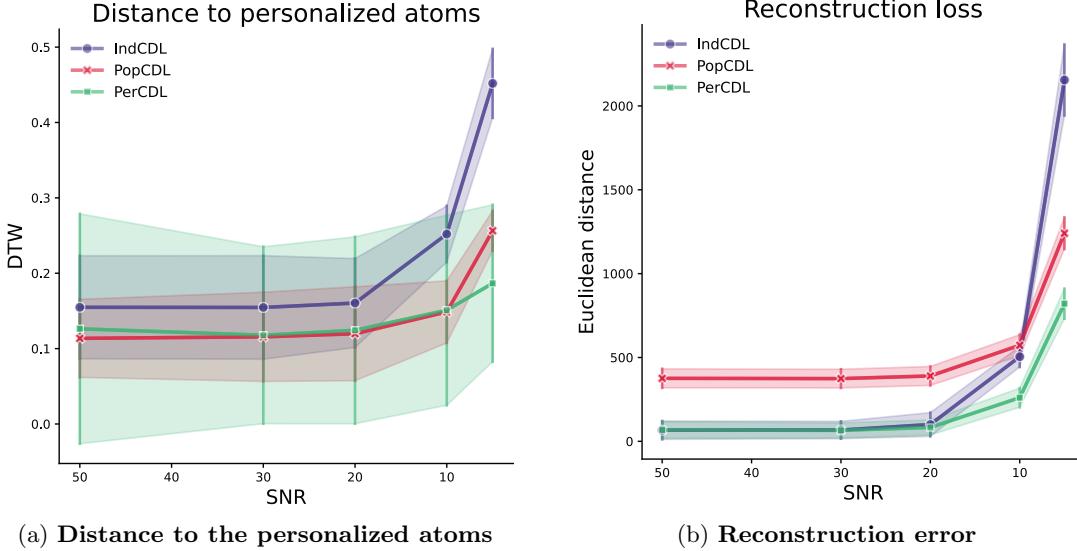


Figure 11: **Robustness of the Different Methods in the Presence of Gaussian Noise.** Personalization hyper-parameters: $D = 3$, $W = 10$.

while wearing an inertial measurement unit (IMU) on the dorsal face of their foot. Two brands of IMUs were employed (XSensTM: 100 Hz, device dimensions $47 \times 30 \times 13$ mm; Technoconcept[®]: 100 Hz, device dimensions $49 \times 38 \times 19$ mm). Participants in the orthopedic group suffered either from lower limb osteoarthritis or from cruciate ligament injury. The neurological group was composed of four specific pathologies: hemispheric stroke, Parkinson’s disease, toxic peripheral neuropathy and radiation-induced leukoencephalopathy.

During testing, 50 participants from each group were chosen at random. For each, we focused specifically on the angular velocity of the right foot in the sagittal plane, as is customary in gait analysis (Voisard et al., 2024). This resulted in a dataset of size $S = 150$, with $N = 15457$. A single common atom ($K = 1$) was searched for, of size set to $L = 100$ to match the length of the gait cycle. Personalization hyper-parameters were set to $D = 4$, $W = 15$ following a brief fine-tuning process. Optimization parameters were left to default, i.e., $n_{\text{steps}} = 25$ with a step size of 10^{-3} , and $\lambda = 10^{-2}$. The common dictionary was initialized with the gait cycle of a random participant from the Healthy group.

Gait Segmentation. PerCDL’s gait segmentation performance (i.e., the location of its activations) was compared to the state-of-the-art gait segmentation method (Voisard et al., 2024). Briefly, Voisard et al. used a sophisticated template-based matching algorithm to identify gait cycles event in accelerometry signals. To compare the two methods in a fair manner, we assessed whether they could identify the start of a gait cycle, with a tolerance margin set relatively high ($\tau_{\text{tol}} = 100$ samples or roughly one second) as the precise beginning of a gait cycle is an ambiguous concept. The number of true positives, i.e., the number of correctly identified gait cycles, the number of false negatives, i.e., the number of gait cycles that were undetected, and the number of false positives, i.e., the number of wrong detection, were computed. The dataset contained expert annotations which acted as a ground truth. Finally, we observed that PerCDL was sometimes too sensitive to small movements of the body preceding and succeeding locomotion: to prevent bias in the results, activations with an absolute value smaller than $\tau_{\text{act}} = 2\%$ of the average of non-zero values were removed. Results are presented in Table 2.

Table 2: **Segmentation Performance of PerCDL vs. SOTA.** The gait signals of 150 participants were used. True positive rate: rate of correctly identified gait cycles. False negatives rate: rate of undetected gait cycles. Proportion of false positives: number of wrong detection divided by the total number of gait cycles.

| | Voisard et al. | PerCDL |
|-------------------------------|----------------|--------|
| True positives rate | 0.959 | 0.988 |
| False negatives rate | 0.040 | 0.012 |
| Proportion of false positives | 0.004 | 0.033 |

Our findings suggest that PerCDL is competitive with the state-of-the-art in detecting the gait cycles. In the dataset evaluated, it offers a better detection than Voisard et al. (2024), although with a slight increase in the number of false detection. Note that this effect could be improved by fine-tuning the activation threshold $\tau_{act} = 2\%$, which was not done in this simple experiment.

IndCDL vs. PopCDL vs. PerCDL. This section compares the three methods on the gait dataset. Initialization was identical for IndCDL, PopCDL and PerCDL. Normalized reconstruction errors for the different methods are as follows: IndCDL = 0.055, PopCDL = 0.109, PerCDL = 0.063, IndCDL barycenter = 0.116. IndCDL’s barycenter corresponds to the reconstruction of the dataset with the Euclidean barycenter of IndCDL’s results.

Because IndCDL essentially over-fits each signal, it presents the smallest reconstruction error. However, it fails to capture the global structure, as a reconstruction of the dataset with the averaged atom identified with IndCDL yields the highest reconstruction error (referred to as IndCDL’s barycenter). By construction, PopCDL captures the shared structure but lacks reconstruction power. Finally, PerCDL builds from the shared structured to fit each signal specifically, thus maintaining both global and local information, i.e., harnessing the structure representation capabilities from PopCDL while reaching reconstruction results similar to IndCDL.

In Figure 12, we observe that the average of IndCDL individual atoms and the common atoms of PopCDL are similar.

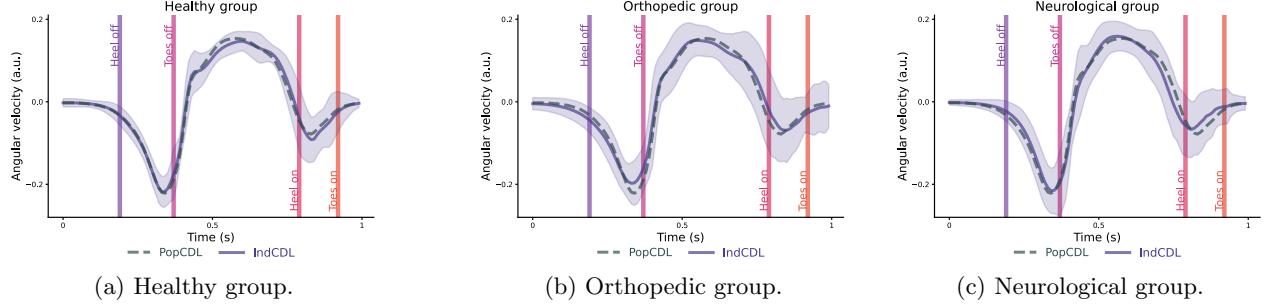


Figure 12: **Comparison of IndCDL and PopCDL on the Healthy, Orthopedic and Neurological Groups.** The shaded region corresponds to 2 standard deviations of the individual atoms returned by IndCDL, and the vertical colored lines timestamp the different phases of the gait cycle. The solid lines of IndCDL are its Euclidean barycenter on a given group (Healthy, Orthopedic, Neurological).

IndCDL. Let us try to conduct a similar analysis as the one presented in Section 3.2 with IndCDL. Initialization was identical to that of PerCDL. Figure 13 presents the group variability analysis of the atoms identified.

A striking result is the presence of “parasitic variability” at the start and end of each atom. This variability is uninformative of the gait cycle, and simply corresponds to an overfit of each signal caused by IndCDL’s learning scheme. This observation suggests that PerCDL is better suited than IndCDL for this kind of analysis.

Group-level analysis with PopCDL. When explicit groups can be identified in a dataset, one may be tempted to apply CDL to each group independently. This reduces to applying PopCDL to each sub-population in the data separately and should intuitively yield a result somewhere between IndCDL and PopCDL. To test this idea, we ran PopCDL on each group from the gait dataset (Table 3).

Table 3: **Group-level Analysis with PopCDL.** The values correspond to the DTW distance between the common atoms from the two groups estimated independently using PopCDL.

| | Healthy - Orthopedic | Healthy - Neurological | Orthopedic - Neurological |
|--------------|----------------------|------------------------|---------------------------|
| DTW distance | 0.0531 | 0.0542 | 0.0367 |

Our results suggest that the common structures found by PopCDL on the different groups are highly similar (similar distance between Healthy vs. Orthopedic and Healthy vs. Neurological, and small distance between

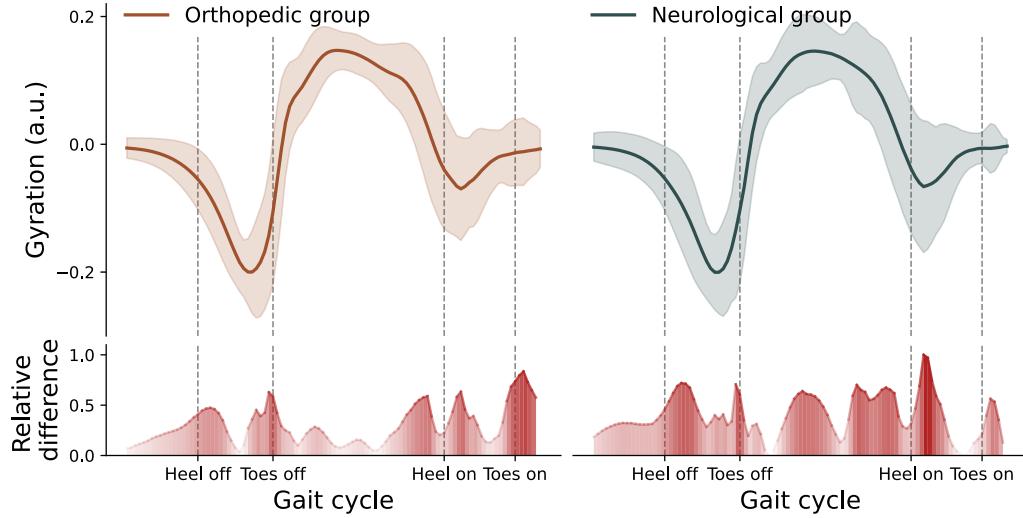


Figure 13: **Application of IndCDL to Locomotion Data.** Atoms learned by PerCDL in the Orthopedic (solid brown line) and Neurological (solid green curve) groups, with two standard deviations (shaded areas). Below each gait cycle, the red-colored curve represents the relative variability in the gait cycle for the current population compared with the Healthy group.

Orthopedic and Neurological). This emphasizes the idea that in the context of physiological, it is often the variation around the common structures that gives interesting information about an individual; something that only PerCDL can capture, not PopCDL.

PerCDL’s Personalized Atoms. The personalized atoms found by PerCDL were investigated. Figure 14 shows an example of the personalized atom found in a single participant per pathology group.

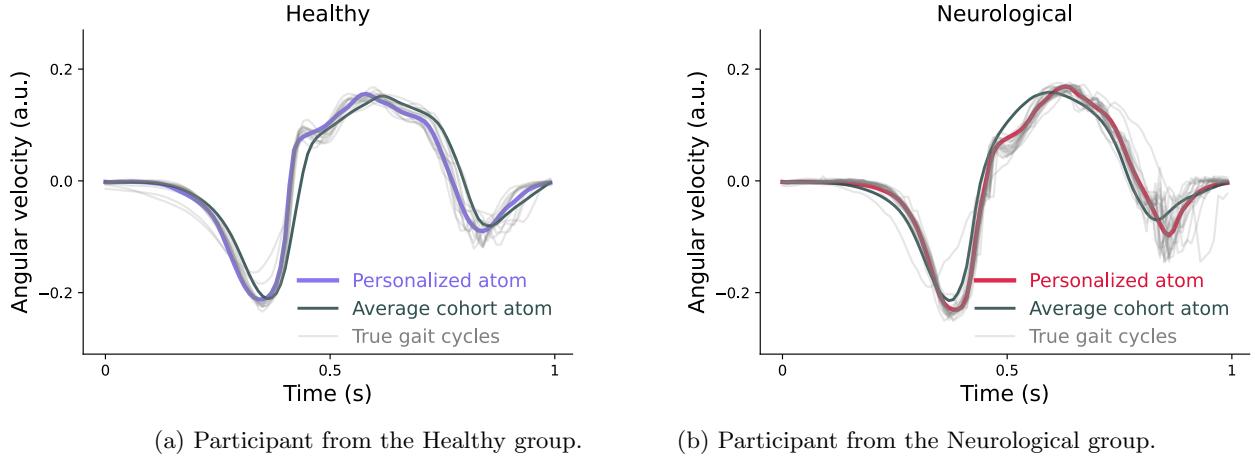


Figure 14: **Example of the Personalized Atom Found in a Single Participant per Pathology Group.** Personalized atoms identified by PerCDL (colored lines) in a single participant, average personalized atom per pathology group (bold black line), and all raw occurrences of the gait cycle in the participant’s signal (light grey lines). PerCDL efficiently captures individual-specific variability.

We observe that the personalized atoms differ from the population’s average and instead accurately match the raw gait cycles manually extracted from an expert’s annotation. Thus, PerCDL is able to efficiently capture participant-specific variability in their gait cycle.

This results was further quantified in Table 4. As expected, the personalized atoms have the smallest distance

to the true gait cycles, compared to the common atoms or the population average. This further demonstrates the efficiency of PerCDL’s personalization step.

Table 4: Euclidean Distance Between the True Gait Cycles and the Atoms. The personalized atom corresponds to the personalized atom learned by PerCDL for the given subject. The cohort personalization mean corresponds to the atom obtained after averaging all personalized atoms for the given pathology group. The distance between the true cycles and the personalized atom is smaller than with all other atoms.

| | Healthy | Orthopedic | Neurological |
|---|---------|------------|--------------|
| True cycles vs. Personalized atom | 0.160 | 0.143 | 0.184 |
| True cycles vs. Cohort personalization mean | 0.305 | 0.274 | 0.302 |
| True cycles vs. Common atom (PerCDL) | 0.299 | 0.350 | 0.330 |
| True cycles vs. Common atom (PopCDL) | 0.192 | 0.481 | 0.554 |

C.6 Application to ECG Data

IndCDL, PopCDL and PerCDL were evaluated on a public ECG dataset⁴ (PTB-XL; Wagner et al., 2020). A subset of $S = 2650$ signals, each of dimension = 12 (due to the 12 ECG leads: $I, II, III, AVL, AVR, AVF, V1 - V6$) and sampled at 100 Hz, were randomly selected.

Prior to the experiments, the hyperparameters for PerCDL were briefly tuned on a separate set of 20 total subjects. The selected parameters set is as follows: $D = 4$, $W = 7$, $n_{steps} = 20$, $K = 1$, $L = 65$. All methods were initialized with activations set to zero and the common dictionary was populated with a single ECG motif from a random subject from the healthy group (superclass “NORM”).

Atoms and Reconstruction. IndCDL, PopCDL and PerCDL were applied to the dataset. Figure 15 presents the atoms identified by IndCDL and PerCDL for the II derivation, while Figure 16 illustrates the reconstruction of three II signals with the different methods for patients in the “Myocardial Infarction” (MI) pathology group.

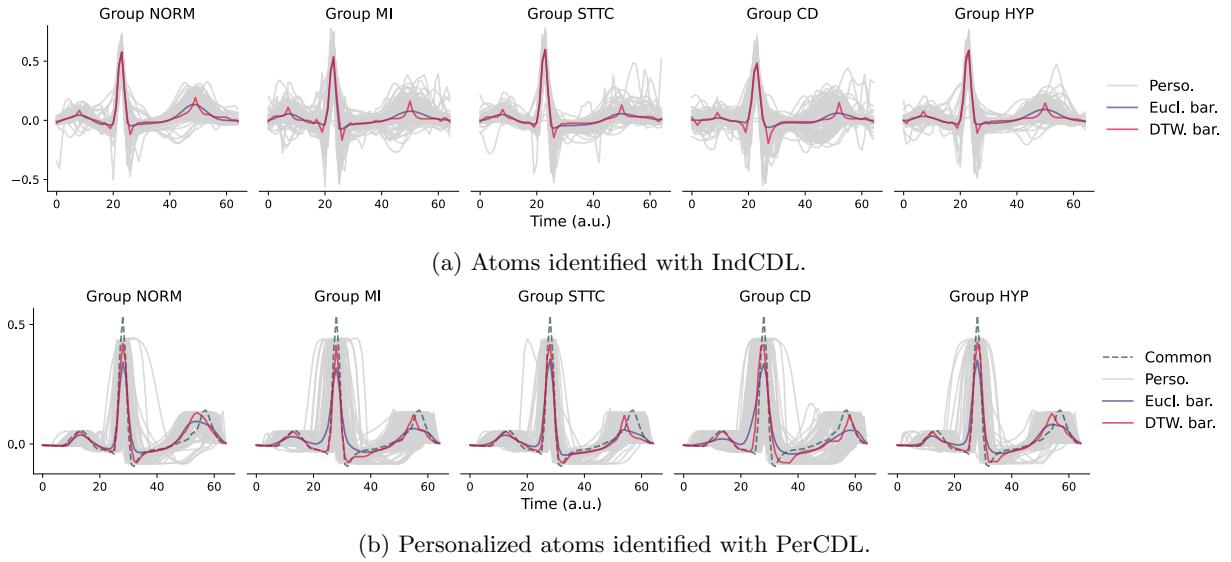


Figure 15: **Atoms Identified by IndCDL and PerCDL in the Different Pathology Superclasses.** (grey) Subject-specific atom. (blue) Euclidean barycenter of all individual atoms. (red) DTW barycenter of all individual atoms.

All methods were able to identify the ECG events efficiently. The dataset presented great variability within each superclass, as they contained multiple specific sub-pathologies. This explains the relatively cluttered plots for the atoms in Figure 15. This effect is strikingly apparent for IndCDL, which has a tendency to overfit each signal, resulting in very “volatile” atoms that may contain noise (see Figure 16, a). Consequently, any effort to average all of the individual atoms is a recipe for failure. On the complete opposite side of the spectrum, PopCDL only learns a common structure that cannot be used directly to model each signal, hence resulting in poor reconstructions. PerCDL constitutes an appealing middle ground, as it captures enough population-level structure to maintain a coherent shape across its atoms, while also capturing enough personal information to reconstruct efficiently each signal. On Figure 16, PerCDL appears to be more robust to noise than IndCDL, although at the price of smoother atoms. Normalized reconstruction errors for the different methods are as follows: IndCDL = 0.176, PopCDL = 0.370, PerCDL = 0.282, IndCDL barycenter = 0.417.

Personalization Parameters. We then wondered whether the personalization parameters learned by PerCDL could efficiently extract meaningful information from the dataset. To that end, we tested whether each signal could be classified into its corresponding pathology superclass (5-class classification task) using only the personalization matrix.

The parameters matrix was separated into train and test sets (70% – 30% split). A simple SVM (*scikit-learn*

⁴The dataset is available online: <https://doi.org/10.13026/kfzx-aw45>.

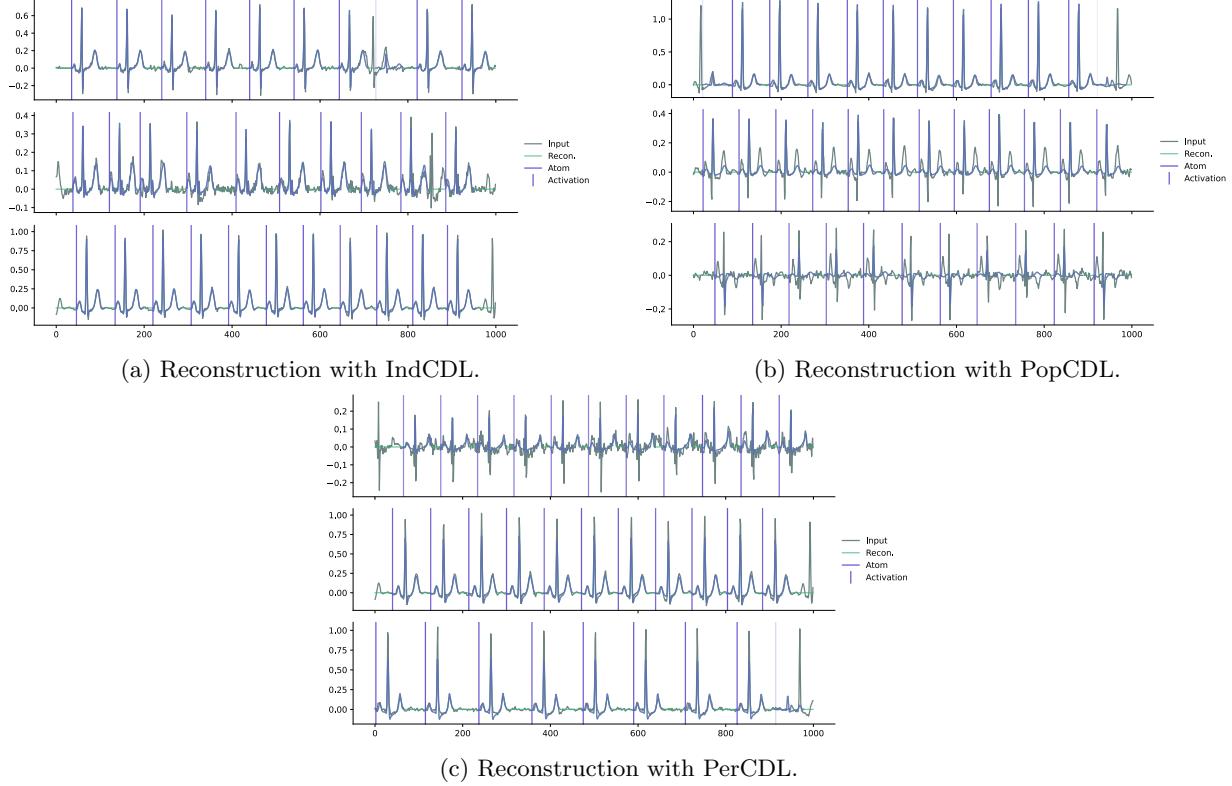


Figure 16: **Reconstruction of Three Example Signals from Three Different Subjects in the MI Group.** (grey) Input signals. (green) Reconstruction. (blue curves) Atoms. (blue, vertical bars) Activations.

Python library, parameters $C = 1$, $\text{gamma}=\text{'scale'}$, $\text{kernel}=\text{'rbf'}$) was then trained on the personalization parameters to classify each signal. We obtained an AUC score (macro, ovo) of 0.864. This result is comparable to the result of some deep learning techniques, namely WAVELET+NN (*cf.*, Strodthoff et al., 2020) that had an AUC score of 0.874.

In order to try and provide a baseline for comparison, we used a k-nearest neighbor classifier (scikit-learn Python library, default parameters) with the raw personalization patterns as inputs. This method obtained an AUC score of 0.728 with the Euclidean distance metric and 0.699 for the DTW distance metric. Thus, PerCDL’s personalization parameters offer a better classification performance than using the personalized atoms directly. Although the primary goal of our paper is not classification-related, our findings suggest nonetheless that the temporal warps employed in PerCDL effectively capture pathology-specific features.

All in all, our findings suggest that the personalization parameters discover relevant characteristics of the input signals.