

SI.1 THEORETICAL BACKGROUND FOR THE CHANNEL SELECTION

1 We describe the background material needed for the proposed two-channel fECG algorithm. The algorithm
 2 is composed of three essential components. The first component is estimating the maternal cardiac activity
 3 in the aECG (maECG) by applying the dsSTFT [1], the beat tracking and the nonlocal median [2], and get a
 4 rough fECG from any given linear combination of the two provided aECG signals. The second component
 5 is the channel selection by applying the lag map [3, 4, 5] and diffusion map (DM) [6] for the sake of
 6 determining the best linear combination of the two channels, which lead to the optimal rough fECG. The
 7 third component is getting the fECG and fetal R peaks information from the optimal rough fECG by again
 8 the dsSTFT and the beat tracking.

9 SI.1.1 Linear combination

10 The main motivation behind the algorithm is motivated by the physiological knowledge of the ECG signal
 11 that among all linear combinations of two channels, with a high probability we could find a combination
 12 that is optimal for the fetal ECG extraction.

Before describing the linear combination idea, recall the well-know *vectocardiogram* (VCG) and its relationship with the ECG signals. It has been well known that the ECG signal, denoted as a continuous time series $E : [0, T] \rightarrow \mathbb{R}$, where $T > 0$ is the observation time, is the projection of the representative dipole current of the electrophysiological cardiac activity on a predesigned direction [7]. Denote the dipole current as a three dimensional continuous time series $d : [0, T] \rightarrow \mathbb{R}^3$. If we could record $d(t)$, it is called the VCG signal. Physiologically, for a normal subject, $d(t)$ is oscillatory with the period $\tau > 0$, which is about 1 second, in the sense that $d(t) \sim d(t + \tau)$ for all $t \in [0, T - \tau]$. Suppose $t_l, l = 1, \dots, m$, where m is the number of cardiac cycles over the period $[0, T]$, is the timestamp corresponding to the maximal amplitude point of the l -th cardiac cycle. We call the vector

$$c = \frac{1}{m} \sum_{l=1}^m d(t_l) \quad (\text{S.1})$$

13 the cardiac axis. For a given ECG signal, there is an associated projection direction $v \in \mathbb{R}^3$ so that E is
 14 the projection of $d(t)$ on v ; that is, $E(t) = v^T d(t)$. It has been well known that depending on v , we could
 15 acquire different aspects of the cardiac information. We mention that in general, v changes according to
 16 time due to the cardiac axis deviation caused by the respiratory activity and other physical movements. To
 17 simplify the discussion, we do not take these facts into account.

Denote d_m to be the mother's VCG and d_f to be the fetus' VCG. Denote c_m to be the mother's cardiac axis and c_f to be the fetus' cardiac axis. Fix two abdominal lead placements and record two aECG signals, denoted as x_1 and x_2 . Denote $v_{m,i} \in \mathbb{R}^3$ and $v_{f,i} \in \mathbb{R}^3$ to be the projection directions of the mother's VCG and fetus' VCG corresponding to x_i , where $i = 1, 2$. Obviously, we have $x_i = v_{m,i}^T d_m + v_{f,i}^T d_f$, where $i = 1, 2$, and it is possible that the fetal cardiac activity is relatively weak in both x_1 and x_2 . To resolve this problem, we consider the following linear combination scheme. Take a linear combination of x_1 and x_2 by

$$\begin{aligned} x_\theta &= \theta x_1 + \sqrt{1 - \theta^2} x_2 & (\text{S.2}) \\ &= \theta v_{m,1}^T d_m + \theta v_{f,1}^T d_f + \sqrt{1 - \theta^2} v_{m,2}^T d_m + \sqrt{1 - \theta^2} v_{f,2}^T d_f \\ &= [\theta v_{m,1}^T + \sqrt{1 - \theta^2} v_{m,2}^T] d_m + [\theta v_{f,1}^T + \sqrt{1 - \theta^2} v_{f,2}^T] d_f, \end{aligned}$$

18 where $\theta \in (-1, 1]$. If these two abdominal leads are placed on two different locations, we know $v_{f,1} \neq v_{f,2}$
 19 and $v_{m,1} \neq v_{m,2}$, and hence the set

$$A := \{\theta v_{f,1} - \sqrt{1 - \theta^2} v_{f,2}\}_{\theta \in (-1,1]} \quad (\text{S.3})$$

20 contains all linear combinations of $v_{f,1}$ and $v_{f,2}$ if we do not distinguish $\theta v_{f,1} - \sqrt{1 - \theta^2} v_{f,2}$ and $-\theta v_{f,1} +$
 21 $\sqrt{1 - \theta^2} v_{f,2}$. Note that the set $\{(\theta, \sqrt{1 - \theta^2})\}_{\theta \in (-1,1]}$ is the 1-dimensional real projective space (identifying
 22 antipodal points of the unit circle, $S^1 := \{x \in \mathbb{R}^2 \mid \|x\| = 1\}$, embedded in \mathbb{R}^2), and it topologically
 23 equivalent to the unit circle S^1 , which is an one dimensional manifold. Based on the above-mentioned
 24 relationship between ECG and VCG, although the fetus could rotate inside the uterus, we know that unless
 25 the cardiac axis c_f of the fetal cardiac activity is perpendicular or almost perpendicular to both $v_{f,1}$ and
 26 $v_{f,2}$, we could find an θ so that x_θ contains a strong fetal cardiac activity. Since the chance that the fetal
 27 cardiac axis is perpendicular to the 2-dim affine subspace corresponding to the two chosen abdominal leads
 28 is low, we could thus conclude that the chance that we could obtain a good signal with strong fECG via the
 29 linear combination scheme is high.

30 SI.1.2 Lag map

31 The lag map is a well-known method widely applied to study a given time series, and its theoretical
 32 foundation has been well established in [3, 8, 9]. In brief, it allows us to reconstruct the structure underlying
 33 the time series. For a given time series \mathbf{f} of length $N \in \mathbb{N}$, the lag map is a mapping from \mathbf{f} to a set of
 34 L -dim points, where L is chosen by the user, via

$$\Psi_{\mathbf{f},L} : i \mapsto (\mathbf{f}(i), \dots, \mathbf{f}(i + L))^T \in \mathbb{R}^{L+1}, \quad (\text{S.4})$$

35 where $i = 1, \dots, N - L$ and the superscript T means taking the transpose. The map $\Psi_{\mathbf{f},L}$ is called the
 36 L -step lag map. It has been shown in [3] that if \mathbf{f} is an observation of a dynamical process whose trajectory
 37 is supported on a d dimensional manifold and L is large enough, then under some weak mathematical
 38 conditions, $\Psi_{\mathbf{f},L}$ could recover the manifold up to a diffeomorphism. Since the cardiac activity is periodic,
 39 the corresponding ‘‘underlying manifold’’ is a one-dimensional circle representing the cardiac dynamics
 40 that is diffeomorphic to the unit circle S^1 , and the lag map of the cardiac activity time series leads to a
 41 point cloud supported on another one-dimensional simple closed curve.

42 The above-mentioned important property of the lag map allows up to examine the quality of the
 43 reconstructed fECG. If $\mathbf{f} \in \mathbb{R}^N$ is the true fECG signal, or a good estimation of the fECG signal,
 44 we obtain an one-dimensional simple closed curve by the point cloud $\mathcal{X}_{\mathbf{f},L} := \{\Psi_{\mathbf{f},L}(i)\}_{i=1}^{N-L} \in \mathbb{R}^{L+1}$.
 45 On the other hand, if the tempted fECG estimator $\mathbf{f} \in \mathbb{R}^N$ fails to be a good estimator of the fECG signal,
 46 the point cloud $\mathcal{X}_{\mathbf{f},L}$ might be away from any one-dimensional simple closed curve. Another important
 47 fact is that when \mathbf{f} is the fECG signal, the point cloud $\mathcal{X}_{\mathbf{f},L}$ is in general non-uniformly sampled from the
 48 one-dimensional circle. This fact comes from the diffeomorphic relationship between the reconstructed
 49 simple closed curve and the underlying simple closed curve via the lag map.

50 SI.1.3 Graph Laplacian and diffusion map

51 To take this important fact into account to examine the quality of the reconstructed fECG via the L -step
 52 lag map, we apply the graph Laplacian (GL), which is the building block of several dimension reduction
 53 algorithms, like the diffusion map (DM) [6, 10]. For a general introduction of GL and DM, we refer readers
 54 to [6, 11, 10]. Here we only provide the necessary steps for our purpose. Fix the given embedded point

55 cloud $\mathcal{X}_{f,L}$. Build a complete affinity graph $G = (V, E, \omega)$ with vertices $V = \mathcal{X}_{f,L}$ by viewing any pairs
 56 of points in $\mathcal{X}_{f,L}$ as edges; that is $E = \{(\Psi_{f,L}(i), \Psi_{f,L}(j)) \mid i \neq j\}$. The affinity function $\omega : E \rightarrow \mathbb{R}^+$ is
 57 then defined by

$$\omega(\Psi_{f,L}(i), \Psi_{f,L}(j)) = \exp \left\{ - \frac{\|\Psi_{f,L}(i) - \Psi_{f,L}(j)\|^2}{\epsilon} \right\}, \quad (\text{S.5})$$

58 for $i, j = 1, \dots, N - L, i \neq j$, and $\epsilon > 0$ is the kernel bandwidth chosen by the user. Here, the affinity
 59 between $\Psi_{f,L}(i)$ and $\Psi_{f,L}(j)$ is reversely proportional to the distance between $\Psi_{f,L}(i)$ and $\Psi_{f,L}(j)$. Note
 60 that while we could choose a more general kernel, here we focus on the Gaussian kernel to simplify
 61 the discussion. We mention that in practice the Gaussian kernel performs well and the dependence on
 62 the chosen kernel is marginal. In general, the point cloud might not be uniformly sampled from the
 63 geometric object we have interest, and the nonuniform sampling effect might generate a negative impact
 64 on the upcoming analysis. To resolve this issue, the α -normalization technique is introduced in [6]. Take
 65 $0 \leq \alpha \leq 1$, we could define an α -normalized affinity function defined on E , denoted as $\omega^{(\alpha)}$, by

$$\omega^{(\alpha)}(\Psi_{f,L}(i), \Psi_{f,L}(j)) = \frac{\omega(\Psi_{f,L}(i), \Psi_{f,L}(j))}{d_i^\alpha d_j^\alpha}, \quad (\text{S.6})$$

66 where d is the degree function defined on the vertex set as

$$d_i = \sum_{j=1}^{N-L} \omega^{(\alpha)}(\Psi_{f,L}(i), \Psi_{f,L}(j)), \quad (\text{S.7})$$

67 for $i = 1, \dots, N - L$. As is shown in [6, 10], when $\alpha = 1$, this α -normalized affinity could effectively
 68 alleviate the impacts introduced by the nonuniform sampling. In our fECG application, as discussed above,
 69 $\mathcal{X}_{f,L}$ is in general non-uniformly sampled from the one-dimensional simple closed curve, so we apply this
 70 α -normalization technique.

71 We are now ready to define the GL. Define an α -normalized affinity matrix $W^{(\alpha)} \in \mathbb{R}^{(N-L) \times (N-L)}$ by

$$W_{ij}^{(\alpha)} := \omega^{(\alpha)}(\Psi_{f,L}(i), \Psi_{f,L}(j)), \quad (\text{S.8})$$

72 for $i, j = 1, \dots, N - L$, define a diagonal α -normalized degree matrix $D^{(\alpha)} \in \mathbb{R}^{(N-L) \times (N-L)}$ by

$$D_{ii}^{(\alpha)} = \sum_{j=1}^{N-L} W_{ij}^{(\alpha)}, \quad (\text{S.9})$$

73 for $i = 1, \dots, N - L$. and the α -normalized graph Laplacian is then defined by

$$L^{(\alpha)} := I - D^{(\alpha)^{-1}} W^{(\alpha)}. \quad (\text{S.10})$$

74 Since L is similar to the symmetric matrix $I - D^{(\alpha)^{-1/2}} W^{(\alpha)} D^{(\alpha)^{-1/2}$, it has a complete set of right
 75 eigenvectors $\phi_1, \dots, \phi_{N-L}$ with corresponding eigenvalues $0 = \lambda_1 < \lambda_2 \leq \dots \leq \lambda_{N-L} \leq 1$. Note
 76 that $\phi_1 = (1, 1, \dots, 1)^T$ since $D^{(\alpha)^{-1}} W^{(\alpha)}$ is a transition matrix defined on the graph G . It has been
 77 shown in [6, 10] that if \mathcal{X} is sampled from a low dimensional Riemannian manifold, when $\alpha = 1$ and
 78 $N \rightarrow \infty$, asymptotically the eigenvectors ϕ_i converges pointwisely and spectrally to the i -th eigenfunction
 79 of the Laplace-Beltrami operator of the Riemannian manifold. In general, this allows us to reconstruct

80 the manifold by applying the diffusion geometry and the spectral embedding theory, which is commonly
81 known as the DM algorithm [6]. The robustness of the GL and DM has been studied in [12, 11].

82 In our problem, due to the periodic oscillation intrinsic to the fECG we have interest, the α -normalized
83 graph Laplacian associated with $\mathcal{X}_{f,L}$ gives us the Laplace-Beltrami operator over a simple closed curve.
84 It follows that asymptotically, the first two non-trivial eigenvectors are the sine and cosine functions. We
85 could thus take this fact into account and design the signal quality index for the channel selection purpose.

REFERENCES

- 86 1 .Lin CY, Li S, Wu HT. Wave-shape function analysis—when cepstrum meets time-frequency analysis.
87 *Journal of Fourier Analysis and Applications* **accepted for publication** (2016).
- 88 2 .Su L, Wu HT. Extract fetal ECG from single-lead abdominal ECG by de-shape short time Fourier
89 transform and nonlocal median. *Frontiers in Applied Mathematics and Statistics* **3** (2017) 2. doi:10.
90 3389/fams.2017.00002.
- 91 3 .Takens F. Detecting strange attractors in turbulence. Rand D, Young LS, editors, *Dynamical Systems*
92 *and Turbulence* (Springer Berlin Heidelberg), *Lecture Notes in Mathematics*, vol. 898 (1981), 366–381.
- 93 4 .Richter M, Schreiber T, Kaplan D. Fetal ECG extraction with nonlinear state-space projections. *IEEE*
94 *Trans. Biomed. Eng.* **45** (1998) 133–137.
- 95 5 .Kotas M, Jezewski J, Matonia A, Kupka T. Towards noise immune detection of fetal QRS complexes.
96 *Comput Methods Programs Biomed* **97** (2010) 241–256.
- 97 6 .Coifman RR, Lafon S. Diffusion maps. *Appl. Comput. Harmon. Anal.* **21** (2006) 5–30.
- 98 7 .Keener J. *Mathematical Physiology* (Springer) (1998).
- 99 8 .Stark J, Broomhead D, Davies M, Huke J. Takens embedding theorems for forced and stochastic
100 systems. *Nonlinear Analysis: Theory, Methods & Applications* **30** (1997) 5303–5314.
- 101 9 .Stark J, Broomhead D, Davies M, Huke J. Delay Embeddings for Forced Systems. II. Stochastic
102 Forcing. *Journal of Nonlinear Science* **13** (2003) 519–577.
- 103 10 .Singer A, Wu HT. Spectral convergence of the connection laplacian from random samples. *Information*
104 *and Inference: A Journal of the IMA* **in press**, doi: **10.1093/imaiai/iaw016** (2016).
- 105 11 .El Karoui N, Wu HT. Connection graph Laplacian methods can be made robust to noise. *Ann. Stat.* **44**
106 (2016) 346–372.
- 107 12 .El Karoui N. On information plus noise kernel random matrices. *The Annals of Statistics* **38** (2010)
108 3191–3216.