ECG Inpainting with denoising diffusion prior.

Lisa Bedin*
Centre de Mathématiques appliquées,
Ecole polytechnique,
lisa.bedin@polytechnique.edu

Gabriel Cardoso*
Centre de Mathématiques appliquées,
Ecole polytechnique,
IHU Liryc, fondation Bordeaux Université,
Univ Bordeaux, CRCTB U4045, INSERM,
gabriel.victorino-cardoso@polytechnique.edu

Remi Dubois
IHU Liryc, fondation Bordeaux Université,
Univ Bordeaux, CRCTB U4045, INSERM,
remi.dubois@ihu-liryc.fr

Eric Moulines
Centre de Mathématiques appliquées,
Ecole polytechnique.
eric.moulines@polytechnique.edu

Abstract

In this work, we train a generative denoising diffusion model (DDGM) in healthy electrocardiogram (ECG) data capable of generating realistic healthy heartbeats. We then show how recent advances in solving linear inverse Bayesian problems with DDGM can be used to derive interpretable outlier detection tools for electrophysiological anomalies.

1 Introduction

Approximately 10% of adult deaths in Europe and the United States are due to sudden cardiac death (SCD), often incorrectly referred to as "cardiac arrest". SCD typically occurs due to extremely rapid ventricular arrhythmias, ie, ventricular fibrillation or ventricular tachycardia (VF/VT). These rapid ventricular arrhythmias are often associated with structural heart disease such as cardiomyopathies or areas of cardiac electrical heterogeneity.

Detecting and quantifying these abnormal heart rhythms with noninvasive techniques such as the electrocardiogram (ECG) is one of the greatest challenges in cardiology. Effective treatments are available to protect at-risk individuals, so accurate assessment is critical. To date, cardiology has relied on left ventricular ejection fraction (LVEF) measurements to assess SCD risk. LVEF, although valuable, has limited utility in younger patients without cardiomyopathies.

Because SCD requires an exceptionally rapid response to prevent deaths, it is extremely difficult to collect noninvasive data directly from this population. An alternative approach is to use the distribution of healthy signals, because databases containing such data are more readily available [14, 34]. This approach may involve detecting outliers or anomalies in the data.

Over the past decade, several techniques have been developed to design and train generative models capable of generating highly realistic patterns from the original data, even for complex high-dimensional

*These authors contributed equally to this work

1st Workshop on Deep Generative Models for Health at NeurIPS 2023.

data types such as images and audio [18, 19, 11]. Denoising-Diffusion Generative Models (DDGMs) have emerged as particularly effective generative models that convert noise to the original data domain through a sequence of denoising steps. These models achieve impressive results in generating images and audio data, and do so without the complexity of adversarial training [27, 31, 28, 29, 5].

Several DDGM trained on ECG have been proposed, but as far as we are concerned, they either try to model a non-centered window of the ECG (instead of a window around an individual heart beat) [2, 20] and/or a given lead [1]. In this work, we focus on modeling the morphology and the inter-lead dependence of the ECG heart beat. To do so, we center the data around the R-peaks to reduce the positional variability of the target distribution, thus making the task of generative model easy. This also considerably simplifies the formulation of morphology-related inverse problems, such as conditional generation of the T-wave from the QRS for example.

Recently, the concept of using generative models as informative priors for solving inverse Bayes problems has attracted considerable attention, mainly due to their flexibility and expressive capabilities [3, 33, 32, 13, 26, 36, 24]. DDGMs are particularly well suited for such tasks, and several studies have investigated their use as priors for solving inverse problems [28, 8, 30, 16, 17]. In particular, recent work by [7] and [35] has proposed consistent algorithms for sampling from the posterior of inverse problems using DDGMs as priors.

Contributions:

- We present the first DDGM specifically trained to generate realistic, healthy electrocardiogram (ECG) beats, with a focus on capturing accurate ECG morphology.
- We use this novel DDGM as a prior in solving inverse problems and demonstrate its practical utility in detecting abnormalities in ECG, in a way that is both efficient and interpretable.

2 Background

2.1 DDGM

This section gives a concise overview of the DDGM and the notations used in this paper. We cover only the basic elements and refer to the original work for complete details and derivations [27, 12, 31, 28]. A denoising diffusion model is a generative model consisting of a forward (noise) and a backward (denoising) process. The forward noising process involves sampling a data point $X_0 \sim \mathsf{q}_{\text{data}}$ from the data distribution, which is then converted into a sequence $X_{1:n}$ of recursively corrupted versions of X_0. In reverse denoising, on the other hand, X_n is sampled to X_n according to an easily acquired reference distribution p_{ref} on $\mathbb{R}^{\mathsf{d}_x}$ and X_0 is generated in $\mathbb{R}^{\mathsf{d}_x}$ by a sequence of denoising steps. Following [15], we use the variance exploding (VE) forward noising process whose joint law is given by

$$\mathsf{q}_{0:n}(x_{0:n}) = \mathsf{q}_{\text{data}}(x_0) \prod_{t=1}^{n} q_t(x_t|x_{t-1}), \quad q_t(x_t|x_{t-1}) = \mathcal{N}(x_t; x_{t-1}, \delta_t^2 \, \mathrm{I}_{\mathsf{d}_x}) \,, \tag{1}$$

where $\mathrm{I}_{\mathsf{d}_x}$ is the identity matrix of size d_x, $\{\delta_t^2\}_{t\in\mathbb{N}} \subset (0,1)$ is a non-increasing sequence and $\mathcal{N}(\mathbf{x}; \mu, \Sigma)$ is the p.d.f. of the Gaussian distribution with mean μ and covariance matrix Σ (assumed to be non-singular) evaluated at $\mathbf{x}$. For all $t > 0$, set $\sigma_t^2 = \sum_{\ell=1}^{t} \delta_\ell^2$ with the convention $\sigma_0^2 = 0$. We have $q_t(x_t|x_0) = \mathcal{N}(x_t; x_0, \sigma_t^2 \, \mathrm{I}_{\mathsf{d}_x})$. The backward denoising process with parameter θ is defined by first choosing n such that $\sigma_n^2 \ggg \sigma_{\text{data}}^2$ and then defining the backward Markov chain $\mathsf{p}_{0:n}^\theta(x_{0:n}) = \mathsf{p}_n(x_n) \prod_{t=1}^{n} p_{t-1}^\theta(x_{t-1}|x_t)$ with

$$\mathsf{p}_n(x_n) = \mathcal{N}\big(x_n; 0_{\mathsf{d}_x}, \sigma_n^2 \, \mathrm{I}_{\mathsf{d}_x}\big) \,, \quad p_{t-1}^\theta(x_{t-1}|x_t) = \mathcal{N}\left(x_{t-1}; \tfrac{\sigma_{t-1}^2}{\sigma_t^2} x_t + \tfrac{\delta_t^2}{\sigma_t^2} \mathcal{D}_{0|t}^\theta(x_t), \tfrac{\sigma_{t-1}^2 \delta_t^2}{\sigma_t^2} \, \mathrm{I}_{\mathsf{d}_x}\right) \tag{2}$$

where 0_{d_x} is the null vector of size d_x and $\mathcal{D}_{0|t}^\theta(x_t)$ is a neural network with parameters θ. The parameter θ is obtained (see [28, Theorem 1]) by solving the following optimization problem:

$$\theta_* \in \operatorname{argmin}_\theta \sum_{t=1}^{n} \left(\frac{1}{\sigma_{t-1}^2} - \frac{1}{\sigma_t^2}\right) \int \|x_0 - \mathcal{D}_{0|t}^\theta(x_0 + \sigma_t^2 \epsilon)\|_2^2 \mathcal{N}(\epsilon; 0_{\mathsf{d}_x}, \mathrm{I}_{\mathsf{d}_x}) \mathsf{q}_{\text{data}}(\mathrm{d}x_0) \mathrm{d}\epsilon \,. \tag{3}$$

The time 0 marginal $\mathsf{p}_0^{\theta_*}(x_0) = \int \mathsf{p}_{0:n}^{\theta_*}(x_{0:n})\mathrm{d}x_{1:n}$ which we will refer to as the *prior* is used as an approximation of $\mathsf{q}_{\mathrm{data}}$ and the time s marginal is $\mathsf{p}_s^{\theta_*}(x_s) = \int \mathsf{p}_{0:n}^{\theta_*}(x_{0:n})\mathrm{d}x_{1:s-1}\mathrm{d}x_{s+1:n}$. In the rest of the paper we drop the dependence on the parameter θ_*.

2.2 Bayesian Inpainting and `MCGdiff`

In the Bayesian inpainting problem, we observe a part of state $y \in \mathbb{R}^{\mathsf{d}_\mathsf{y}}$, which can be written as $y = (x[i])_{i\in J}$, where $J \subset [1 : \mathsf{d}_x]$ is a set of d_y distinct observed state indexes and $x[i] \in \mathbb{R}$ is the i-th coordinate of the vector $x \in \mathbb{R}^{\mathsf{d}_x}$. For $x \in \mathbb{R}^{\mathsf{d}_x}$ we denote by $\overline{x} \in \mathbb{R}^{\mathsf{d}_\mathsf{y}}$ the coordinates in J and we denote J_c the set of free (unobserved) indices and $\underline{x} \in \mathbb{R}^{\mathsf{d}_x - \mathsf{d}_\mathsf{y}}$ the vector of free coordinates. The posterior density defined only for the free coordinates $\underline{x}$ is $\phi_0^y(\underline{x}) \propto \mathsf{p}_0(y^\frown\underline{x})$, where $^\frown$ is the reconstruction operator based on the indices J, J_c, i.e., for all $x \in \mathbb{R}^{\mathsf{d}_x}$, $\overline{x}^\frown\underline{x} = x$. To obtain approximate samples from $\phi_0^y(\underline{x})$, we use the `MCGdiff` algorithm introduced in [7]. `MCGdiff` is a Sequential Monte Carlo (SMC) sampler that targets the sequence of distributions $\{q_t(\overline{x}_t|y)\mathsf{p}_t(x_t)\}_{\in[1:n]}$ with a finite set of N particles. One of the main advantages of `MCGdiff` is that it is consistent, meaning that the Kullback-Leibler distance between `MCGdiff` samples and the target posterior goes to 0 with a ratio of N^{-1}.

3 Numerical Examination

3.1 Dataset

The dataset includes the training set used in the 2021 PhysioNet Challenge as indicated in [10, 22, 23]. For each patient in the dataset, records were assigned to one of three groups, training, Cross-validation (CV), and test, with probabilities of (0.8, 0.1, 0.1), respectively.

The preprocessing procedure includes four main phases: downsampling to a frequency of 250 Hz, extraction of R peaks, beat extraction, and subsequent normalization. More details on the preprocessing pipeline can be found in the Appendix.

3.2 DDGM

Our goal is to develop a generative model capable of generating healthy ECG beats, focusing on QRS and T-wave morphology. Because the (I, II, III) standard ECG leads and the augmented leads (aVL, aVR, aVF) are linearly correlated, we decided to build a generative model exclusively for the (aVL, aVR, aVF, V1, V2, V3, V4, V5, V6) leads. We denote the ECG signal, which serves as our state variable, as $X \in \mathbb{R}^{176\times 9}$.

Factors such as age (A), sex (S), and preceding R-R interval (RR) are known to influence QRS and T-wave morphology, as evidenced by the existing literature (see [21, 25, 4]). Therefore, we train a conditional DDGM that learns to generate the random variable $X|A, S, RR$. The training dataset includes all patient records in the training group, with labels representing normal sinus rhythm (NSR), sinus bradycardia (SB), sinus tachycardia (STach), or sinus arrhythmia (SA).

We have a comparable architecture and training strategy to the one in "EDM" [15], except that we replaced the original 2D convolutional layers with 1D convolutional layers. In addition, we introduced a positional embedding for the temporal axis within the data, and this embedding was treated similarly to the noise level positional embedding throughout the network. Comprehensive details on the architecture and training hyperparameters can be found in the Appendix. In each configuration examined, the model that achieved the lowest training loss in the cross-validation group (CV) was saved. As far as we know, we are the first to propose a generative model for healthy heartbeats, which makes the validation of such a model a real challenge. To establish an appropriate validation metric that evaluates how accurately the generative model reproduces data that resemble the distribution of the training data, we rely on the Earth Mover Distance

(EMD) [*]. EMD is an optimal transport distance metric that can be computed from samples only. In Figure 4, we present the EMD values between generated samples and samples extracted from both the training and test datasets, varying the number of steps for the backward generation process. The generation process is conditioned on the same population as the test data (2757 signals), and the conditioning variables are factored into the EMD calculation. The sampled ECGs are shown in the appendix fig. 5. We also include the EMD between the test and and same-size batches from the training dataset as a reference (the red dashed line indicates the mean over multiple batches). For an insight into the EMD's magnitude, we additionally illustrate the EMD between a corrupted test dataset with a noise distribution of $\mathcal{N}(0, 2.5 \times 10^{-3}\,\mathrm{I})$ and the training dataset (depicted in green).

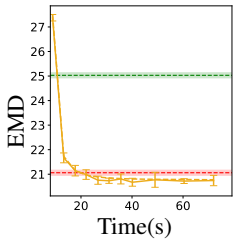

Figure 1: Generated ECGs using the DDMG from section 3.2.

3.3 Abnormality detection

In this section, we demonstrate how a DDGM for healthy heart beats, as introduced in Section 3.2, and the `MCGdiff` when applied to the in-painting problem, can be employed as a transparent anomaly detection tool. We consider the set of observed coordinates J to correspond to the augmented limb leads (aVL, aVR, aVF). Those leads imparts insights into the heart's general orientation and overall geometry. Those leads are also further away from the heart than the pre-cordial leads (V1, V2, V3, V4, V5, V6) and measure a resultant of the electrical activity of the whole heart. Therefore, we might expect that the presence of localised (small) electrical anomalies to be mainly in the pre-cordial leads and not perceptible on the augmented limb leads. That's the reason we attempt to reconstruct the pre-cordial leads from the augmented limb leads using `MCGdiff` and compare the distance between the posterior sample with a healthy DDGM as prior and the actual pre-cordial leads of a given patient. We expect patients with localised electrical cardiac anomalies to have a higher distance, which would indicate how far away this patient is from the "normality".

We sample 100 `MCGdiff` samples per patient for 200 patients per label using $N = 50$ particles. The rationale for the choice of $N = 50$ is deferred to the appendix. For each setting, we calculate the Mahalanobis distance between each non conditioned track and the posterior, which is defined as $d_{\mathrm{MH}}(x) := \sqrt{(\underline{x} - \mu)^T \Sigma^{-1} (\underline{x} - \mu)}$ where μ and Σ are estimated using the posterior samples.

We then take the maximum over all the non conditioned tracks. The boxplot of those values per label are given in fig. 2 for the labels MI (Miocardial Infarction), IRBBB (Incomplete Right bundle branch block), RBBB (Right bundle branch block) and LBBB (Left bundle branch block) which all correspond to electrical anomalies during the ventricular activity. We observe that the healthy patients (NSR) show smaller values of the Mahalanobis distance. We show in fig. 3 the posteriors ECG and the patients ECG. This illustrates one of the main advantages of our method, namely the visualisation of the posterior. We can provide precise information on where the signal is different from what is expected by the generative model, something that is not possible in black-box outlier algorithms.

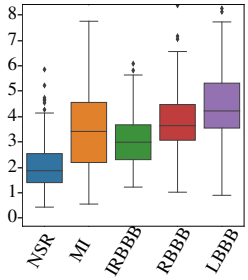

Figure 2: Distribution of maximum Mahalanobis distances for populations from the test set.

[*]We use the POT library [9] for the EMD calculations.

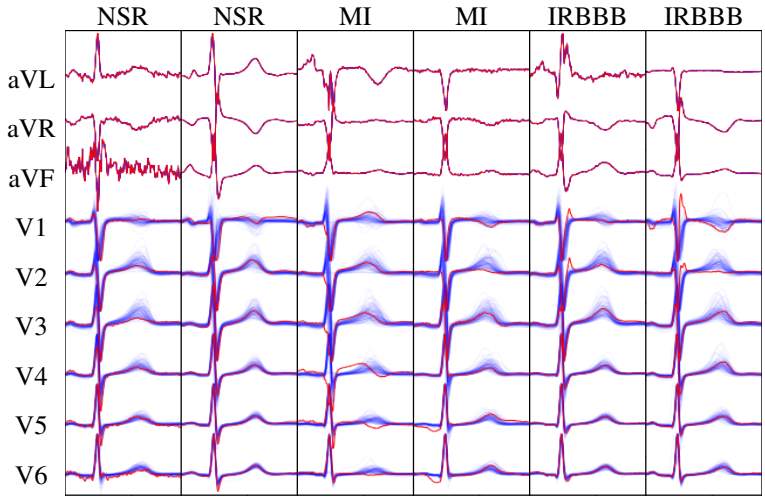

Figure 3: The figure displays both the posterior ECG (in blue) and the patient's ECG (in red) for different labels on the training set.

4 Conclusion

We proposed a DDGM trained on healthy heart ECG signals that is able to accurately reproduce the data. We then show how this can be combined with the recent advances in solving linear inverse problems with DDGM priors to generate an interpretable white-box outlier detection algorithm for the ECG, that allows for the visualization of what is the expected ECG for a given conditioning configuration against the patients ECG.

References

[1] Edmonmd Adib, Amanda S. Fernandez, Fatemeh Afghah, and John J. Prevost. Synthetic ecg signal generation using probabilistic diffusion models. *IEEE Access*, 11:75818–75828, 2023.

[2] Juan Miguel Lopez Alcaraz and Nils Strodthoff. Diffusion-based conditional ecg generation with structured state space models. *Computers in Biology and Medicine*, 163:107115, 2023.

[3] Siavash Arjomand Bigdeli, Matthias Zwicker, Paolo Favaro, and Meiguang Jin. Deep mean-shift priors for image restoration. *Advances in Neural Information Processing Systems*, 30, 2017.

[4] Robyn L Ball, Alan H Feiveson, Todd T Schlegel, Vito Stare, and Alan R Dabney. Predicting "heart age" using electrocardiography. *Journal of personalized medicine*, 4(1):65–78, 2014.

[5] Joe Benton, Yuyang Shi, Valentin De Bortoli, George Deligiannidis, and Arnaud Doucet. From denoising diffusions to denoising markov models. *arXiv preprint arXiv:2211.03595*, 2022.

[6] Jan C. Brammer. biopeaks: a graphical user interface for feature extraction from heart- and breathing biosignals. *Journal of Open Source Software*, 5(54):2621, 2020.

[7] Gabriel Cardoso, Yazid Janati El Idrissi, Sylvain Le Corff, and Eric Moulines. Monte carlo guided diffusion for bayesian linear inverse problems, 2023.

[8] Hyungjin Chung, Jeongsol Kim, Michael Thompson Mccann, Marc Louis Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. In *The Eleventh International Conference on Learning Representations*, 2023.

[9] Rémi Flamary, Nicolas Courty, Alexandre Gramfort, Mokhtar Z. Alaya, Aurélie Boisbunon, Stanislas Chambon, Laetitia Chapel, Adrien Corenflos, Kilian Fatras, Nemo Fournier, Léo Gautheron, Nathalie T.H. Gayraud, Hicham Janati, Alain Rakotomamonjy, Ievgen Redko, Antoine Rolet, Antony Schutz, Vivien Seguy, Danica J. Sutherland, Romain Tavenard, Alexander Tong, and Titouan Vayer. Pot: Python optimal transport. *Journal of Machine Learning Research*, 22(78):1–8, 2021.

[10] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. Ch. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220, 2000 (June 13). Circulation Electronic Pages: http://circ.ahajournals.org/content/101/23/e215.full PMID:1085218; doi: 10.1161/01.CIR.101.23.e215.

[11] Jie Gui, Zhenan Sun, Yonggang Wen, Dacheng Tao, and Jieping Ye. A review on generative adversarial networks: Algorithms, theory, and applications. *IEEE transactions on knowledge and data engineering*, 2021.

[12] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33:6840–6851, 2020.

[13] Sebastian Kaltenbach, Paris Perdikaris, and Phaedon-Stelios Koutsourelakis. Semi-supervised invertible neural operators for bayesian inverse problems. *Computational Mechanics*, pages 1–20, 2023.

[14] Jingsu Kang and Hao Wen. A Study on Several Critical Problems on Arrhythmia Detection using Varying-Dimensional Electrocardiography. *Physiological Measurement*, 43(6):064007, 6 2022.

[15] Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. In *Proc. NeurIPS*, 2022.

[16] Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. In *Advances in Neural Information Processing Systems*.

[17] Bahjat Kawar, Gregory Vaksman, and Michael Elad. Snips: Solving noisy inverse problems stochastically. *Advances in Neural Information Processing Systems*, 34:21757–21769, 2021.

[18] Diederik P Kingma, Max Welling, et al. An introduction to variational autoencoders. *Foundations and Trends® in Machine Learning*, 12(4):307–392, 2019.

[19] Ivan Kobyzev, Simon JD Prince, and Marcus A Brubaker. Normalizing flows: An introduction and review of current methods. *IEEE transactions on pattern analysis and machine intelligence*, 43(11):3964–3979, 2020.

[20] Huayu Li, Gregory Ditzler, Janet Roveda, and Ao Li. Descod-ecg: Deep score-based diffusion model for ecg baseline wander and noise removal. *IEEE Journal of Biomedical and Health Informatics*, pages 1–11, 2023.

[21] Marek Malik, Katerina Hnatkova, Donna Kowalski, James J Keirns, and E Marcel van Gelderen. Qt/rr curvatures in healthy subjects: sex differences and covariates. *American Journal of Physiology-Heart and Circulatory Physiology*, 305(12):H1798–H1806, 2013.

[22] Matthew A Reyna, Nadi Sadr, Erick A Perez Alday, Annie Gu, Amit J Shah, Chad Robichaux, Ali Bahrami Rad, Andoni Elola, Salman Seyedi, Sardar Ansari, Hamid Ghanbari, Qiao Li, Ashish Sharma, and Gari D Clifford. Will two do? varying dimensions in electrocardiography: The physionet/computing in cardiology challenge 2021. In *2021 Computing in Cardiology (CinC)*, volume 48, pages 1–4, 2021.

[23] Matthew A. Reyna, Nadi Sadr, Erick A. Perez Alday, Annie (Ping) Gu, Amit J. Shah, Chad Robichaux, Ali Bahrami Rad, Andoni, Elola, Salman Seyedi, Sardar Ansari, Hamid Ghanbari, Qiao, Li, Ashish Sharma, and Gari D. Clifford. Issues in the automated classification of multilead ecgs using heterogeneous labels and populations. *Physiological Measurement*, 43, 2022.

[24] Teemu Sahlström and Tanja Tarvainen. Utilizing variational autoencoders in the bayesian inverse problem of photoacoustic tomography. *SIAM Journal on Imaging Sciences*, 16(1):89–110, 2023.

[25] Guy Salama and Glenna CL Bett. Sex differences in the mechanisms underlying long qt syndrome. *American Journal of Physiology-Heart and Circulatory Physiology*, 307(5):H640–H648, 2014.

[26] Hyomin Shin and Minseok Choi. Physics-informed variational inference for uncertainty quantification of stochastic differential equations. *Journal of Computational Physics*, page 112183, 2023.

[27] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International Conference on Machine Learning*, pages 2256–2265. PMLR, 2015.

[28] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *International Conference on Learning Representations*, 2021.

[29] Yang Song, Conor Durkan, Iain Murray, and Stefano Ermon. Maximum likelihood training of score-based diffusion models. *Advances in Neural Information Processing Systems*, 34:1415–1428, 2021.

[30] Yang Song, Liyue Shen, Lei Xing, and Stefano Ermon. Solving inverse problems in medical imaging with score-based generative models. In *International Conference on Learning Representations*, 2022.

[31] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021.

[32] Jingwen Su, Boyan Xu, and Hujun Yin. A survey of deep learning approaches to image restoration. *Neurocomputing*, 487:46–65, 2022.

[33] Xinyi Wei, Hans van Gorp, Lizeth Gonzalez-Carabarin, Daniel Freedman, Yonina C Eldar, and Ruud JG van Sloun. Deep unfolding with normalizing flow priors for inverse problems. *IEEE Transactions on Signal Processing*, 70:2962–2971, 2022.

[34] Hao Wen and Jingsu Kang. Hybrid Arrhythmia Detection on Varying-Dimensional Electrocardiography: Combining Deep Neural Networks and Clinical Rules. In *2021 Computing in Cardiology (CinC)*. IEEE, 9 2021.

[35] Luhuan Wu, Brian L. Trippe, Christian A. Naesseth, David M. Blei, and John P. Cunningham. Practical and asymptotically exact conditional sampling in diffusion models, 2023.

[36] Xu Zhihang, Xia Yingzhi, and Liao Qifeng. A domain-decomposed vae method for bayesian inverse problems. *arXiv preprint arXiv:2301.05708*, 2023.

A Appendix

A.1 Extension of `MCGdiff` to the Variance Exploding framework.

In [7], they consider the forward process to be the so called variance preserving framework [31]. As shown in [15], the variance exploding framework, coupled with properly chosen discretization of the backward sampler, shown better performance in uncoditional generatation. Therefore, we decide to use this framework in this paper. Applying Bayes formula to the forward process eq. (1), we obtain

$$q_{n|0}(x_n|x_0) = \mathcal{N}\left(x_n; x_0, \sigma_n^2 \mathrm{I}\right), \quad q_{t-1|t,0}(x_{t-1}|x_t, x_0) = \mathcal{N}\left(x_{t-1}; \boldsymbol{\mu}_t(x_0, x_t), \tilde{\sigma}_t^2 \, \mathrm{I}_d\right),$$
$$\boldsymbol{\mu}_t(x_0, x_t) = \frac{\sigma_{t-1}^2 x_t + \delta_t^2 x_0}{\sigma_t^2}, \quad \tilde{\sigma}_t^2 = \frac{\sigma_{t-1}^2}{\sigma_t^2}\delta_t^2 .$$

By following the derivation in [7], we obtain the following proposal kernel and weight functions for the SMC algorithm:

- Proposal Kernel:

$$p_s^y(x_s|x_{s+1}) = \underline{p}_s(\underline{x}_s|x_{s+1})\mathcal{N}\left(\overline{x}_s; \mathsf{k}_s y + (1-\mathsf{k}_s)\boldsymbol{\mu}_{s+1}(\overline{x}_{s+1}, \overline{\boldsymbol{\chi}}_{0|s+1}(x_{s+1})), \sigma_s^2 \mathsf{k}_s\right)$$

  where $\mathsf{k}_s = \frac{\tilde{\sigma}_s^2}{\tilde{\sigma}_s^2 + \sigma_s^2}$.

- The weight function $\widetilde{\omega}_s$ is chosen as follows; $\widetilde{\omega}_0(\overline{x}_1) = \overline{p}_0(y|\overline{x}_1)/\overline{q}_{1|0}(\overline{x}_1|y)$,

$$\begin{aligned}\widetilde{\omega}_{n-1}(x_n) &= \int p_{n-1}(x_{n-1}|x_n)\overline{q}_{n-1|0}(\overline{x}_{n-1}|y)\mathrm{d}x_{n-1} \\ &= \mathcal{N}\left(y; \boldsymbol{\mu}_n(\overline{x}_n, \overline{\mathcal{D}_{0|n}}(x_n)), \tilde{\sigma}_n^2 + \sigma_s^2\right)\end{aligned}$$

  and for $s \in [1 : n-2]$,

$$\begin{aligned}\widetilde{\omega}_s(x_{s+1}) &= \int \overline{p}_s(\overline{x}_s|x_{s+1})\overline{q}_{s|0}(\overline{x}_s|y)\mathrm{d}x_s / \overline{q}_{s+1|0}(\overline{x}_{s+1}|y) \\ &= \frac{\mathcal{N}\left(y; \boldsymbol{\mu}_{s+1}(\overline{x}_{s+1}, \overline{\mathcal{D}_{0|s+1}}(x_{s+1})), \tilde{\sigma}_{s+1}^2 + \sigma_s^2\right)}{\mathcal{N}\left(y; \overline{x}_{s+1}, \sigma_{s+1}^2\right)} .\end{aligned} \tag{4}$$

A.2 Numerics

Preprocessing Pipeline: The Down-sampling phase consists of down-sampling all the signals to a sampling frequency of 250hz. We then proceed to the R-peak extraction phase, where we extract the channel-wise first principal component of the whole ECG. This track then is passed through a Savitzky-Golay filter of order 3 with window length 15. Then, the r-peaks are extracted using the method proposed in [6]. From there, a window is extracted around each r-peak r by selecting all the tracks in the window $[r - 48 : r + 128]$. We then divide each track in the window by its respective maximum absolute value attained during QRS complex.

Number of particles: As the number of particles N increases, the distance between the target posterior distribution and the distribution of the particles from `MCGdiff` decreases. The question of interest is thus when is N big enough. To obtain a initial guess into a good number of N, we selected a patient in the test dataset and draw 10^3 samples from `MCGdiff` with $N = 10^4$. We consider this samples as the reference samples, thus as the posterior samples. We then generated for several different N, 10^3 samples using `MCGdiff` and calculated the EMD distance w.r.t the reference samples. Figure 4 shows the evolution of the EMD with the number of particles N and we see that $N = 50$ seems to yield a good trade-off between inference time and accuracy.

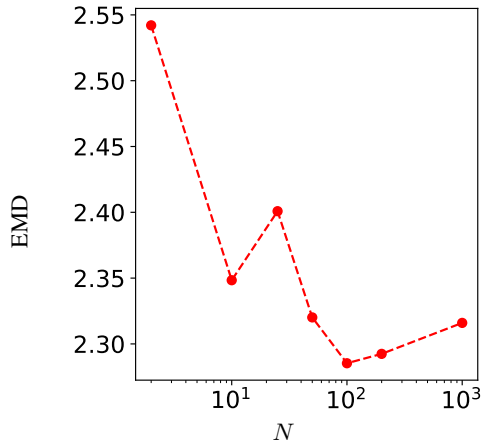

Figure 4: EMD distance between 1000 samples from `MCGdiff` with N particles and 1000 samples of `MCGdiff` with 10^5 particles, that is considered the standard samples.

Generated ECGs: We display in fig. 5 some generated ECGs from the DDGM. The conditioning features are taken randomly from the test dataset.

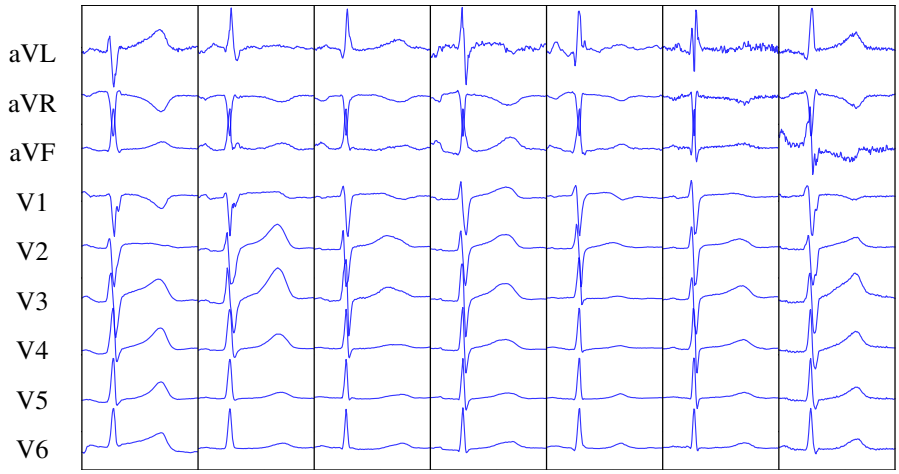

Figure 5: ECGs generated from the DDGM with random conditioning features sampled from the test dataset.

Architecture details: We implement a very close architecture to [15] and available at `https://github.com/NVlabs/edm` as well as training procedure. The main difference is that we replaced the 2D convolutional layers by 1D ones in every UNet. The final network use the following parameters:

- First embedding dimension: 192,
- Number of Unet blocks per resolution: 2,
- Number of resolutions: 1,
- Dropout probability 0.10.

For the training, the following configuration was used:

- learning rate: 10^{-4},
- Number of epochs: 10^4,
- Batch Size: 1024,

- Exponential moving average coefficient: 0.9999.

For the (forward diffusion) we used the following parameters:

- $\sigma_{\min} = 2 \times 10^{-4}$,
- $\sigma_{\max} = 80$,
- $\sigma_{\text{data}} = 0.5$,
- Importance law of σ for training: $\text{Log}\,\mathcal{N}(-1.2, 1.2^2\,\text{I})$.