J. On pair-Exploiting Diffusion Model

In this appendix, we provide more detailed explanations about the training and inference of HEGGS for the clarification.

J.1. Remark:Prediction targets of diffusion model

Referring equation (2),(4),(6) and (7) of (Ho et al., 2020), the forward process $q(x_{1:T}; X)$ would be given by

$$q(x_t|x_{t-1}) = \mathcal{N}(\sqrt{1-\beta_t}x_{t-1}, \beta_t I), q(x_t|x_0) = \mathcal{N}(\sqrt{\overline{\alpha}_t}x_0, (1-\overline{\alpha}_t)I). \tag{24}$$

and thus we have

$$x_t = \sqrt{\overline{\alpha}_t}x_0 + \sqrt{1-\overline{\alpha}_t}\epsilon \text{ where } \epsilon \sim \mathcal{N}(0,1). \tag{25}$$

The backward process $q(x_{t-1}|x_t, x_0)$ would be

$$q(x_{t-1}|x_t, x_0) \sim \mathcal{N}(\tilde{\mu}(x_t, x_0), \tilde{\beta}_t I) \tag{26}$$

where

$$\tilde{\mu}_t(x_t, x_0) = \frac{\sqrt{\overline{\alpha}_{t-1}}\beta_t}{1-\overline{\alpha}_t}x_0 + \frac{\sqrt{\alpha_t}(1-\overline{\alpha}_{t-1})}{1-\overline{\alpha}_t}x_t \text{ and } \tilde{\beta}_t = \frac{1-\overline{\alpha}_{t-1}}{1-\overline{\alpha}_t}\beta_t. \tag{27}$$

In implementation, it is required to find $\tilde{\mu}_t(x_t, x_0)$ term in Equation (27). There are several methods for the prediction, with replacing x_0 by estimate $x_\theta(x_t, t)$. The HEGGS directly predicts x in sample space, as it would be more natural since we want to learn the morphology between paired data, compared to the alternatives which predicts the noise ϵ (Ho et al., 2020) or $\mathbf{v}$-prediction(Salimans & Ho, 2022).

Algorithm 1 HEGGS training

Input: Seismic dataset $\mathbb{D}$, diffusion steps T
repeat
  $(W^{src}, W^{tgt}, \vec{c}_{tgt}) \sim \mathbb{D}$
  convert (W^{src}, W^{tgt}) to (X^{src}, X^{tgt})
  $t \sim Uniform(1, \cdots, T)$
  $\epsilon \sim \mathcal{N}(0,1)$
  Take gradient descent step on
    $\nabla\|X^{tgt} - \mathcal{D}_{AE}(\mathbf{m}_\theta(z_t^{src}, \vec{c}_{tgt}, t))\|^2$
    where $z_t^{src} = \sqrt{\overline{\alpha}_t}\mathcal{E}_{AE}(X^{src}) + \sqrt{1-\overline{\alpha}_t}\epsilon$
until converged

Algorithm 2 Generation

Input: Diffusion steps T,condition vector $\vec{c}_{tgt}$, source waveform W^{src} (optional)
if W^{src} is given **then**
  convert W^{src} to spectrogram X^{src}
  $z_T = \mathcal{E}_{AE}(X^{src})$
else
  sample $z_T \sim \mathcal{N}(0,1)$
end if
for $t = T, \cdots, 1$ **do**
  sample $\mathbf{z} \sim \mathcal{N}(0,1)$
  compute $\tilde{z} = \mathbf{m}_\theta(z_t, \vec{c}_{tgt}, t)$
  compute $z_{t-1} = \tilde{\mu}_t(z_t, \tilde{z}) + \sqrt{\tilde{\beta}_t}\mathbf{z}$ (Eq. 27)
end for
$X^{tgt} = \mathcal{D}_{AE}(z_0)$
Convert X^{tgt} to waveform W^{tgt}
Return: W^{tgt}

J.2. Training with pairs

As described in Section 3, we consider the paired data (X^{src}, X^{tgt}) with corresponding condition vector $\vec{c}_{src}$ and $\vec{c}_{tgt}$. Note that $\vec{c}_{src}$ is not in use.

Since X^{src} and X^{tgt} are the observations of same earthquake, we make assumption that there exist a morphology η which maps the latent x_t^{src} of X^{src} at time t, to x_t^{tgt} using $\vec{c}_{tgt}$, as a random variable. We formulate this assumption with Equation (1), as follows:

$$\eta(x_t^{src}, \vec{c}_{tgt}, t) \sim q(x_t^{tgt}|X^{tgt}) \tag{1}$$

This assumption includes the intuition that the broadband waveform signal is a combination of earthquake information, which is considered to be included in X^{src}, and local geological features near observatory, encoded by positional information from $\vec{c}_{tgt}$.

For training, we aim to train the neural network $\mathbf{m}_\theta$ which is a composition of η and denoising model $\mathbf{x}_\theta$. Precisely, $\mathbf{m}_\theta$ would be written by

$$\mathbf{m}_\theta(x, \vec{c}, t) = \mathbf{x}_\theta(\eta(x, \vec{c}, t), \vec{c}, t). \tag{28}$$

Since $\eta(x_t^{src}, \vec{c}_{tgt}, t) = x_t^{tgt}$, we have $\mathbf{m}_\theta(x_t^{src}, \vec{c}_{tgt}, t) = \mathbf{x}_\theta(x_t^{tgt}, \vec{c}_{tgt}, t)$ for the paired latents (x_t^{src}, x_t^{tgt}), the loss function of diffusion model Equation (2) would be equivalent to Equation (3):

$$\mathcal{L}'_{DM} = \mathbb{E}_{(X^{src}, X^{tgt}, \vec{c}_{tgt}), \epsilon, t} \| X^{tgt} - \mathbf{m}_\theta(x_t^{src}, \vec{c}_{tgt}, t) \|^2 \tag{29}$$

After that, we consider same procedure in latent space (the z_t^{tgt} for the clarification) with autoencoder consist of the encoder $\mathcal{E}_{AE}$ and decoder $\mathcal{D}_{AE}$, we obtain the loss function Equation (9), with end-to-end training.

$$\mathcal{L}_{ours} := \mathbb{E}_{(X^{src}, X^{tgt}, \vec{c}_{tgt}), \epsilon, t} \| X^{tgt} - \mathcal{D}_{AE}(\mathbf{m}_\theta(z_t^{src}, \vec{c}_{tgt}, t)) \|^2 \tag{9}$$

In Algorithm 1, we present an training algorithm for the HEGGS training with $\mathcal{L}_{ours}$. The paired waveforms and corresponding condition vector of target waveform would be sampled from the dataset, and the gradient descent would update all modules $\mathbf{m}_\theta, \mathcal{E}_{AE}$ and $\mathcal{D}_{AE}$ together.

Remark J.1. For the training process of diffusion model with Equation (9), several details below are considered for the loss and model design.

1. During the training, the noise is designed to be added to the Z^{src} instead of Z^{tgt}. This would provide robustness against site-specific noise which already included in observation W^{src} and its latent vector Z^{src}.
2. When t is small, z_t^{src} would be almost same to Z^{src} (this is also because X^{src} itself is already noisy) and thus the model would learn the transformation η with more attention.
3. Regarding the intuition that z_t^{src} and z_t^{tgt} will be identified (in distribution) when t is sufficiently large, the training loss Equation (9) would be equivalent to the conventional training loss for $\mathbf{x}_\theta$ training when we disregard the end-to-end training. Hence, the model learns to generate from the noise w/o W^{src} too, during the training.
4. Since η and $\mathbf{m}_\theta$ does not take $\vec{c}_{src}$ as input. Therefore the model learns to extract common information from z_t^{src} through multiple pairs of observations of same earthquake during training, regardless the local information (encoded by location) of observatory. This makes the model can handle z_t^{tgt} as a input too, since it shares the information of earthquake.

J.3. Inference w/o W^{src}

Although the diffusion model is trained with paired data and takes W^{src} as an input, our model is capable to synthesize seismic waveform without the observation W^{src}.

Since η is defined to map the source latent z_t^{src} to target latent z_t^{tgt}, it also maps the target latent to itself, in distribution. Precisely, we can write

$$\eta(z_t^{tgt}, \vec{c}_{tgt}, t) = z_t^{tgt} \tag{30}$$

and thus the output of neural network would be

$$\mathbf{m}_\theta(z_t^{tgt}, \vec{c}_{tgt}, t) = \mathbf{x}_\theta(\eta(z_t^{tgt}, \vec{c}_{tgt}, t), \vec{c}_{tgt}, t) = \mathbf{x}_\theta(z_t^{tgt}, \vec{c}_{tgt}, t) \tag{31}$$

Therefore, we can use conventional reverse process

$$z_{t-1}^{tgt} = \tilde{\mu}_t(z_t^{tgt}, \mathbf{m}_\theta(z_t^{tgt}, \vec{c}_{tgt}, t)) + \sigma_t \mathbf{z}, \mathbf{z} \sim N(0, I) \tag{32}$$

even if z_T^{tgt} is the gaussian noise sampled from $\mathcal{N}(0, 1)$.

In Algorithm 2, we summarize the generation process of our model. Note that the diffusion steps are equivalent to LDM(Rombach et al., 2022) when W^{src} is not given.