Householder Projector for Unsupervised Latent Semantics Discovery

Yue Song¹, Jichao Zhang¹, Nicu Sebe¹, and Wei Wang²

¹Department of Information Engineering and Computer Science, University of Trento, Italy
²Beijing Jiaotong University, China
yue.song@unitn.it

Abstract

Generative Adversarial Networks (GANs), especially the recent style-based generators (StyleGANs), have versatile semantics in the structured latent space. Latent semantics discovery methods emerge to move around the latent code such that only one factor varies during the traversal. Recently, an unsupervised method proposed a promising direction to directly use the eigenvectors of the projection matrix that maps latent codes to features as the interpretable directions. However, one overlooked fact is that the projection matrix is non-orthogonal and the number of eigenvectors is too large. The non-orthogonality would entangle semantic attributes in the top few eigenvectors, and the large dimensionality might result in meaningless variations among the directions even if the matrix is orthogonal. To avoid these issues, we propose Householder Projector, a flexible and general low-rank orthogonal matrix representation based on Householder transformations, to parameterize the projection matrix. The orthogonality guarantees that the eigenvectors correspond to disentangled interpretable semantics, while the low-rank property encourages that each identified direction has meaningful variations. We integrate our projector into pre-trained StyleGAN2/StyleGAN3 and evaluate the models on several benchmarks. Within only 1% of the original training steps for fine-tuning, our projector helps StyleGANs to discover more disentangled and precise semantic attributes without sacrificing image fidelity.

Code is publicly available via https://github.com/KingJamesSong/HouseholderGAN.

1. Introduction

Generative Adversarial Networks (GANs) [15] and the recent style-based generative models (StyleGANs) [27, 28, 26] in particular have become the leading paradigm of generative modeling in the vision domain. The latent spaces of StyleGANs are known to embed rich and hierarchical semantics [14, 21]: moving the latent code in certain directions could trigger meaningful variations in the output images. Therefore, latent semantics discovery methods

![Figure 1: Motivation of our proposed Householder Projector. Here “Projector” denotes the projection matrix that maps latent codes to features, i.e., the modulation weight of StyleGANs. (Top Left) The singular value imbalance of the non-orthogonal projector would entangle multiple semantics in the top interpretable directions. (Top Right) Due to the large dimensionality of the projector, directly enforcing vanilla orthogonality would spread the data variations among all the eigenvectors, leading to imperceptible and meaningless traversal. (Bottom) Our Householder Projector equips the projection matrix with low-rank orthogonal properties, which simultaneously disentangles semantics into multiple equally-important eigenvectors and guarantees that each direction could correspond to semantically-meaningful variations.]
emerge to identify such interpretable directions that each variation factor is disentangled and the generation process can be precisely controlled [14, 47, 42, 61, 48, 1, 58, 37, 8].

Among the recent explorations of unsupervised interpretable semantics discovery methods [56, 48, 57, 41], SeFa [48] pointed out a promising direction to discover semantically meaningful concepts by computing the eigenvector of the projector. Here we refer to the projection matrix that maps latent codes to features as the projector. The key observation is that using the eigenvectors of the projector for latent perturbation would maximize the data variations. Such identified eigenvectors/directions would correspond to meaningful semantic concepts. However, as shown in Fig. 1 top left, this is likely to cause semantics entangled in the top few eigenvectors. This phenomenon stems from the fact that the variation caused by an eigenvector is actually determined by the associated eigenvalue. Due to the imbalanced eigenvalue distribution, the discovered directions are not equally-important, and the top few ones would simultaneously manipulate multiple attributes. This eigenvalue discrepancy can be mitigated by enforcing orthogonality constraint to the projector. Nonetheless, since standard orthogonal matrices have as many equally-important eigenvectors as the dimensionality, there would not be enough semantics to mine in practice when the method scales to large models such as StyleGANs whose projector dimension is 1024 or 512. Consequently, as an accompanying limitation, the data variations would be split among all the eigenvectors and none of them could produce meaningful output changes (see Fig. 1 top right).

To resolve the above issues, we propose Householder Projector, a low-rank orthogonal matrix representation based on Householder transformations, to parameterize the projection matrix that maps latent codes to features. The projector is first decomposed to its SVD form (USV$^T$). Next, the orthogonal singular vectors U and V are represented by a series of Householder reflectors, respectively. Thanks to the normalization of Householder reflections, the orthogonality is also preserved during backpropagation. For the singular value S, we explicitly set it as a low-rank identity matrix (i.e., $S = \text{diag}(1, \ldots, 1, 0, \ldots, 0)$) whose rank defines exactly the number of semantic concepts. As shown in Fig. 1 bottom, the low-rank property guarantees that the identified directions would cause meaningful variations, while the orthogonality encourages that each semantic attribute is disentangled from the others. Moreover, a proper initialization scheme is proposed to leverage the statistics of pre-trained weights, and an acceleration technique is applied to speed up the computation. We also propose a metric dedicated to measuring the smoothness of latent space to interpretable directions based perturbations. Our Householder Projector is integrated into pre-trained StyleGANs [28, 26] at multiple different layers to mine the diverse and hierarchical semantics. Since our projector inevitably changes the pre-trained parameters, the modified models incorporated with our projector are fine-tuned for limited steps to maintain the original image fidelity. Both quantitative and qualitative results on several widely used benchmarks [29, 63, 10, 25, 13] show that within marginal fine-tuning steps (1% of the training steps), our Householder Projector improves the latent semantics discovery of StyleGANs to have more precise attribute control while not impairing the quality of generated images.

The key results and main contributions are as follows:

- We propose Householder Projector, a flexible and general low-rank orthogonal matrix representation based on Householder transformations, to parameterize the projector of generative models for latent semantics discovery.
- Our Householder Projector can be easily integrated into pre-trained GAN models. With the acceleration technique, it can be efficiently fine-tuned for very limited additional steps, which paves the way for applying latent semantics discovery and orthogonality techniques to large-scale generative models such as StyleGANs.
- Extensive experiments on two popular backbones (StyleGAN2 [28] and StyleGAN3 [26]) and six benchmarks (FFHQ [29], LSUN Church and Cat [63], AFHQ [10], and SHHQ [13]) demonstrate that within marginal extra fine-tuning steps (1% of the original training steps), our method could both achieve precise attribute control and preserve the original image quality.

2. Related Work

Generative Adversarial Networks. In the past few years, GAN-based generative models [15] have achieved remarkable progress in high-fidelity image synthesis [43, 24, 5, 27, 28, 26, 25, 39, 6, 16, 50, 46]. The generation process usually takes the following procedure: a randomly-sampled latent code is fed into the generator through a projection step, and then the generator outputs realistic-looking images. Recently, the style-based generators [27, 28, 26] that gradually absorb layer-wise latent style codes are becoming the de facto GAN backbones. StyleGAN2 [28] improves the original StyleGAN [27] by redesigning generator normalization and training techniques, and StyleGAN3 [26] further explores some equivariance properties. We use the popular StyleGAN2 and StyleGAN3 as our backbones in this paper.

Latent Semantics Discovery. Recently, lots of methods explore disentangling the latent space to achieve image editing by moving the latent code in the identified interpretable directions [7, 3, 21, 47, 17, 56, 67, 48, 18, 41, 54, 57, 55, 30, 55, 31, 60, 22, 44, 9, 53, 51]. Supervised methods rely on human annotations (i.e., segmentation masks, attribute categories, 3D priors, and text descriptions) to de-
fine the semantic labels $[14, 47, 61, 21, 42, 33, 8, 12, 49, 40, 34, 62, 59, 23]$. Here we mainly highlight some relevant unsupervised approaches that are free of annotations. Voynov et al. [56] proposed to jointly learn a set of directions and a classifier such that the interpretable directions can be recognized by the classifier. More recently, Peebles et al. [41] and Wei et al. [57] proposed to add orthogonal Hessian and Jacobian regularization to encourage disentangled representations, respectively. Song et al. [51] proposed to use wave equations to model the spatiotemporal dynamic non-linear latent traversals in generative models. SeFa [48] showed that the eigenvectors of the projector after the latent code could maximize the data variations and proposed to directly use them as the interpretable directions. However, due to the imbalanced eigenvalues, the imbalanced eigenvalues would cause different extents of variations due to the imbalance of associated eigenvalues. Supposing that $\mathbf{n}$ is an eigenvector of $\mathbf{A}^T \mathbf{A}$, then we would have $||\mathbf{A} \mathbf{n}||_2^2 = \sigma^2$ where $\sigma$ is the corresponding singular value of $\mathbf{A}$. For non-orthogonal matrices, the singular values are exponentially decreasing, i.e., $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_d$. This would cause most variations captured by the first few interpretable directions. The imbalance is thus likely to make semantic attributes entangled in the top eigenvectors (see Fig. 1 top left).

3.2. Householder Low-rank Orthogonal Projector

The eigenvalue discrepancy can be eliminated by enforcing strict orthogonality. Orthogonal matrices have the property of identical eigenvalues, which assigns equal importance to each eigenvector. The low-rank constraint could further limit the number of semantics to mine. Therefore, we propose to use Householder representation, a flexible and general framework to parameterize matrices, to endow the projection matrix with low-rank orthogonality.

**Householder Parameterization.** Householder representations can parameterize any matrices by using a series of Householder reflectors to represent the orthogonal singular vectors of its Singular Value Decomposition (SVD) form. In the field of deep learning, it has been used to parameterize the transition matrix and to stabilize gradients of recurrent neural networks [36, 64, 35]. The key to the orthogonality representation relies on the following theorem:

**Theorem 1** (Householder representation [20, 32]). Given any square orthogonal matrix $\mathbf{M} \in \mathbb{R}^{m \times m}$, it can be represented by the product of Householder matrices $\mathbf{M} = \mathbf{H}_1 \mathbf{H}_2 \ldots \mathbf{H}_m$ where each Householder matrix is parameterized by a vector as $\mathbf{H}_i = I - 2 \frac{\mathbf{h}_i \mathbf{h}_i^T}{||\mathbf{h}_i||_2^2}$.

Let $\text{USV}^T$ denote the SVD of the projector $\mathbf{A}$ where $\mathbf{S}$...
Figure 2: Gallery of some semantic attributes discovered by our Householder Projector across all used datasets (FFHQ [29] in the top row, LSUN Church [63] and LSUN Cat [63] in the 2nd row, SHHQ [13] in the 3rd row, and MetFaces [25] and AFHQ [10] in the bottom row). These semantic attributes are sorted from low-level layers (left) to high-level layers (right).
denotes the diagonal singular value, and \( U \) and \( V \) are left and right orthogonal singular vectors. We use the accumulation of Householder reflectors (i.e., \( \prod_{i=0}^{h} (I - 2 \frac{u_i v_i^T}{\|u_i\|^2}) \)) and \( \prod_{i=0}^{w} (I - 2 \frac{u_i v_i^T}{\|v_i\|^2}) \) where \( w, h \) denotes the width and height of \( A \) to parameterize \( U \) and \( V \), respectively. Notice that only \( u_i \) and \( v_i \) are the actual learnable parameters. For \( S \), we explicitly set it to a diagonal matrix and keep the weight fixed. The projector is thus represented by our Householder parameterizations.

**Low-rank Constraint.** To achieve the property of disentangled attributes, one straightforward approach is to parameterize \( A \) as an orthogonal matrix, i.e., to set \( S \) to an identity matrix \( I \) where \( I_{i,j} = 1 \) for \( i = j \) and \( I_{i,j} = 0 \) otherwise. This would lead to equally-important semantic attributes whose number is exactly the projector dimension. However, for large generative models such as StyleGANs, the projector dimension is typically very large (e.g., 512 or 1024). It is not likely to have enough attributes to edit in practice. Setting \( S \) to a full-rank diagonal matrix would spread data variations among all the eigenvectors, resulting in trivial and imperceptible traversal (see Fig. 1 top right). To avoid this issue, we propose to define \( S \) as a low-rank identity matrix:

\[
S = \text{diag}(1, \ldots, 1, 0, \ldots, 0),
\]

where \( N \) defines the rank and also the number of semantic attributes to mine. By restricting the rank of the projector, we explicitly limit the number of interpretable directions. As shown in Fig. 1 bottom, this would benefit the latent traversal for more meaningful output variations.

**Orthogonality Preservation.** One advantage of our Householder representation is that the orthogonality can be kept during backpropagation. Since we have the vector normalization \( \frac{h_i}{\|h_i\|^2} \), the impact of any gradient descent step \( h_i - \eta \nabla h_i \) on the orthogonality would be cancelled, i.e., the updated vector remains orthogonal after normalization.

**Initialization.** When our method is applied to pre-trained models, the well-trained network weights could be leveraged to initialize our Householder Projector. To this end, we propose to use the nearest-orthogonal mapping [52] to project the weight matrix into its orthogonal form that has the nearest distance in the Frobenius norm (i.e., \( \min \| R - A \|_F \)) where \( R^T R = I \). The closed-form solution is given by \( U V^T \) where \( U S V^T \) is the SVD of the original weight matrix \( A \). Next, we decompose \( U \) and \( V \) into their Householder reflectors and use them to initialize our projector. Such an initialization scheme leverages the statistics of the original weight matrix, which might give our projector a good starting point and improve the performance (see the ablation study of Sec. D.3 in the supplementary).

**Acceleration.** The accumulated product of elementary

![Figure 3: Illustration on how our Householder Projector](image)

Figure 3: Illustration on how our Householder Projector modifies the original StyleGAN architecture. The projector is parameterized by its SVD form where \( U \) and \( V \) are represented by the accumulation of Householder reflectors, and \( S \) is set to a low-rank identity matrix. Our projector is applied at multiple different layers of StyleGANs to explore the diverse and hierarchical semantics. The actual learnable parameters are \( u_i \) and \( v_i \).

Householder matrices can be accelerated via the theorem:

**Theorem 2** (Compact WY representation [4]). For any accumulation of \( m \) Householder matrices \( H_1 \ldots H_m \), there exists \( W, Y \in \mathbb{R}^{d \times m} \) such that \( I - 2 W Y^T = H_1 \ldots H_m \) where computing \( W \) and \( Y \) takes \( O(d m^2) \) time and \( m \) sequential Householder multiplications.

This theorem indicates the possibility to repeatedly split the accumulation \( H_1 H_2 \ldots H_m \) into multiple sub-sequences until irreducible. Then these sub-sequences can be computed in parallel and gradually merged. As revealed in the ablation of Sec. D.3 in the supplementary, this technique could fully exploit the parallel computational power of GPUs and greatly speed up the aggregation, particularly in our case where the projector dimension is large.

**Implementation in StyleGANs.** Fig. 3 depicts how our Householder Projector modifies the original StyleGAN architectures. The projector used in the weight modulation module is represented by our proposed projector. The weight matrix is thus endowed with low-rank orthogonal properties. We insert the proposed projector at every layer of StyleGAN2 and every four layers of StyleGAN3. Since StyleGAN3 has 15 intermediate layers, the adjacent layers have very similar and even repeated semantics. Therefore,
4. Experiments

In this section, we first introduce the experimental setup, followed by the quantitative and qualitative evaluation. We defer the full ablation studies to Sec. D of supplementary.

4.1. Setup

Models. We evaluate our Householder Projector on StyleGAN2 [28] and StyleGAN3 [26], i.e., the challenging state-of-the-art GAN backbones in the field of computer vision.

Datasets and Baselines. For StyleGAN2, we conduct experiments on FFHQ [29], LSUN Church [63], and LSUN Cat [63]. The experiments of StyleGAN3 are performed on SHHQ [13], MetFaces [25] and AFHQv2 [10]. We mainly compare our method with representative unsupervised latent semantics discovery approaches, including SeFa [48], Orthogonal Jacobian Regularization (OrJaR) [57], and Hessian Penalty (HP) [41]. SeFa [48] can be directly applied to pre-trained models, while OrJaR and HP need extra fine-tuning or training from scratch due to the regularization.

Metrics. We conduct quantitative evaluation using (1) Fréchet Inception Distance (FID) [19], (2) Perceptual Path Length (PPL) [27], (3) Perceptual Interpretable Path Length (PIPL), and (4) Face Attribute Correlation. FID aims to measure the image quality and diversity by computing the distance between the real and fake distributions, and PPL is designed to assess the perceptual smoothness of the latent space where the smoothness can reflect the disentanglement ability. Our proposed PIPL is a natural modification of PPL by adapting the latent manipulation from random interpolation-based perturbations to vector-based perturbations using interpretable directions. Compared with PPL, our PIPL can better measure the latent space smoothness when the latent code is perturbed along with specific vectors, which suits more vector-based semantic discovery methods like SeFa [48] and ours. Furthermore, for StyleGAN2 trained on FFHQ, we rely on pre-trained face attribute estimators to compute the correlation coefficient between the traversal steps and the face attributes. Besides the above four metrics, we also assess the disentanglement through visual observation. We defer the details of used datasets and metrics to Sec. C of the supplementary.

Implementation Details. We adopt the widely used PyTorch implementation of StyleGAN2\(^1\) and convert the official TensorFlow pre-trained models into PyTorch for FFHQ [29], LSUN Church [63], and LSUN Cat [63]. For StyleGAN3, we use the official code and pre-trained models of AFHQv2 [10] and MetFaces [25]. As for SHHQ [13], we also use the official pre-trained model\(^2\). Following the original optimization strategy, we finetune all the parameters of the pre-trained generator and discriminator within 1\(\%\) of the total training steps (kimgs for StyleGAN). For instance, the fine-tuning process takes 5\(K\) steps for StyleGAN2 with FFHQ and 250 kimgs for StyleGAN3 with AFHQ. The fine-tuning time is thus very limited due to the small number

\(^1\)https://github.com/rosinality/stylegan2-pytorch
\(^2\)https://github.com/stylegan-human/StyleGAN-Human
of training steps. To give concrete examples, fine-tuning models on FFHQ and AFHQ takes 1.5 and 2.5 hours, respectively. The rank $N$ of $\mathbf{S}$ is set to 10 for all experiments based on our empirical observation of the number of semantics of StyleGANs. Our editing strength is set the same as SeFa. We use 4 RTX A6000 GPUs for the training. For the comparison fairness, the baseline methods OrJaR [57] and HP [41] are also fine-tuned with the same steps.

4.2. Qualitative Evaluation and Discussion

Diverse and Precise Attributes. Fig. 2 exhibits some semantic attributes discovered by our Householder Projector on all the datasets. Our method mines a diverse set of disentangled semantics, enabling a wide range of attribute manipulation. Take as an example the first row of attributes discovered on FFHQ [29]. The left columns present diverse high-level semantic concepts such as “Pose” and “Shape”, while the right columns show low-level attributes such as “Color” and “Lighting”. This hierarchy also meets the same trend of original StyleGANs. The manipulation of the diverse and highly-disentangled semantics would give users more precise control on the image generation process. 

Semantic Unambiguity. Importantly, our interpretable di-
Comparison against Other Methods. Fig. 7 and Fig. 8 compare the latent traversal of some directions against other baselines on FFHQ [29] and SHHQ [13], respectively. All the methods can discover similar attributes in the same layer, such as the head length in Fig. 7 left. However, the baselines suffer from entangled semantics and the other attributes vary during the traversal, such as hairstyle for SeFa [48] and OrJaR [57], and expression for HP [41]. By contrast, our method is able to discover more precise semantics and preserve other non-target attributes.

Comparison of Same Samples. To better compare the qualitative disentanglement performance, we further use a GAN inversion technique (PTI [45]) to create nearly the same images from FFHQ for different methods. Fig. 6 displays some qualitative comparisons. For the same samples, our method still has more precise attribute control.

Local Editing Applications. With precise attribute control, our method is even able to edit local regions by simply performing latent traversal. Fig. 4 displays such a use case. Our method achieves very competitive performance against local editing methods such as the recent ReSeFa [66]. In addition, our method is free of extra bounding box as input.

Please refer to Sec. E of the supplementary for more visualizations about comparisons on other datasets, and semantic diversity/unambiguity/hierarchy.

4.3. Quantitative Evaluation

StyleGAN2 Results. Table 2 presents the quantitative evaluation results on FFHQ [29], LSUN Church [63] and LSUN Cat [63] datasets. Our proposed Householder Projector outperforms other baselines in terms of both PPL and PIPL. This demonstrates that our method has a smoother and more structured latent space, which corresponds to more disentangled representations. In particular, our approach surpasses SeFa [48] by a large margin, which indicates the benefit of enforcing low-rank orthogonality to the projection matrix. Compared with the soft orthogonality regularization used in OrJaR [57] and HP [41], the hard orthogonality of our projector also has an advantage in latent smoothness.
due to the strict orthogonality preservation and the additional low-rank constraint. Moreover, our FID score is also very competitive, implying that our method could simultaneously improve the disentanglement performance while keeping the quality of generated images unharmed. For the attribute correlation, we also use GAN inversion to create a dataset of $2K$ identical images for each method. Table 1 compares the correlation results on FFHQ. Our method outperforms other unsupervised baselines and preserves the attribute well in particular.

<table>
<thead>
<tr>
<th>Datasets</th>
<th>Methods</th>
<th>FID ((\downarrow))</th>
<th>PPL ((\downarrow))</th>
<th>PIPL ((\downarrow))</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OrJaR [57]</td>
<td>17.61</td>
<td>5171.66</td>
<td>3.700</td>
</tr>
<tr>
<td></td>
<td>HP [41]</td>
<td>17.32</td>
<td>5132.27</td>
<td>4.663</td>
</tr>
<tr>
<td></td>
<td>Ours</td>
<td>16.89</td>
<td>4192.52</td>
<td>0.099</td>
</tr>
<tr>
<td>AFHQv2 [10] 512 × 512</td>
<td>SeFa [48]</td>
<td>4.40</td>
<td>2193.74</td>
<td>0.470</td>
</tr>
<tr>
<td></td>
<td>OrJaR [57]</td>
<td>5.45</td>
<td>2103.47</td>
<td>0.440</td>
</tr>
<tr>
<td></td>
<td>HP [41]</td>
<td>5.33</td>
<td>2198.46</td>
<td>0.463</td>
</tr>
<tr>
<td></td>
<td>Ours</td>
<td>4.98</td>
<td>2052.38</td>
<td>0.070</td>
</tr>
<tr>
<td>SHHQ [13] 512 × 256</td>
<td>SeFa [48]</td>
<td>2.54</td>
<td>1621.07</td>
<td>0.370</td>
</tr>
<tr>
<td></td>
<td>OrJaR [57]</td>
<td>4.78</td>
<td>1614.56</td>
<td>0.245</td>
</tr>
<tr>
<td></td>
<td>HP [41]</td>
<td>5.38</td>
<td>1648.74</td>
<td>0.216</td>
</tr>
<tr>
<td></td>
<td>Ours</td>
<td>4.17</td>
<td>1549.01</td>
<td>0.119</td>
</tr>
</tbody>
</table>

Table 3: Evaluation results on StyleGAN3.

**StyleGAN3 Results.** Table 3 compares the performance of our method against other baselines on MetFaces [25], AFHQv2 [10], and SHHQ [13] datasets. The results are very coherent with those on StyleGAN2: our Householder Projector improves the latent space smoothness without harming the image fidelity. Our method as well as the two baselines that involve fine-tuning have slightly worse FID than the original StyleGAN3. This might stem from the fact that due to the limited computational resources, our used batch size (16) is actually smaller than the original setting of StyleGAN3 (32). As revealed in the ablation study of Sec. D.2 of the supplementary, the hyper-parameter batch size has a substantial impact on FID. We expect that increasing the batch size would further boost the image quality of our method and lead to a more competitive FID score.

## 5. Conclusion

This paper proposes a general and flexible low-rank orthogonal matrix representation coined as Householder Projector for unsupervised latent semantics discovery of generative models. The proposed method endows the projection matrix with low-rank orthogonality. This superiority helps pre-trained GANs to achieve precise and diverse semantics control within limited fine-tuning steps. Extensive experiments of StyleGANs on various benchmarks demonstrate that our method could simultaneously identify the disentangled attributes while maintaining image fidelity.

**Acknowledgement**

This work was supported by the MUR PNRR project FAIR (PE00000013) funded by the NextGenerationEU and by the PRIN project CREATIVE (Prot. 2020ZSL9F9).

**References**


2


[62] Huiting Yang, Liangyu Chai, Qiang Wen, Shuang Zhao, Zixun Sun, and Shengfeng He. Discovering interpretable latent space directions of gans beyond binary attributes. In CVPR, 2021. 3


[64] Jiong Zhang, Qi Lei, and Inderjit Dhillon. Stabilizing gradients for deep neural networks via efficient svd parameterization. In ICML, 2018. 3


[66] Jiapeng Zhu, Yujun Shen, Yinghao Xu, Deli Zhao, and Qifeng Chen. Region-based semantic factorization in gans. ICML, 2022. 3, 5, 8