Dualing GANs

Yujia Li[1,*], Alexander Schwing[3], Kuan-Chieh Wang[1,2] and Richard Zemel[1,2]

[1]University of Toronto [2]Vector Institute
[3]University of Illinois at Urbana-Champaign [*]Now at DeepMind

Instability in GAN Training

- Saddle point formulation for GAN training:

$$\max_{\theta} \min_{\mathbf{w}} f(\theta, \mathbf{w})$$

Instability in GAN Training

- Saddle point formulation for GAN training:

$$\max_{\theta} \min_{\mathbf{w}} f(\theta, \mathbf{w})$$

- Alternate gradient updates for θ and $\mathbf{w}$

$$\theta \to \theta + \eta_\theta \nabla_\theta f(\theta, \mathbf{w}), \quad \mathbf{w} \to \mathbf{w} - \eta_{\mathbf{w}} \nabla_{\mathbf{w}} f(\theta, \mathbf{w}).$$

- This leads to instability.

Proposed solution: $\max_\theta \min_{\mathbf{w}} \rightarrow \max_\theta \max_\lambda$

- Dualize $\min_{\mathbf{w}} f(\theta, \mathbf{w})$ into $\max_\lambda g(\theta, \lambda)$.

- $\max_\theta \max_\lambda g(\theta, \lambda)$ is more stable than $\max_\theta \min_{\mathbf{w}} f(\theta, \mathbf{w})$.

GANs with Linear Discriminator

- Linear scoring function $F(\mathbf{w}, \mathbf{x}) = \mathbf{w}^\top \mathbf{x}$, discriminator

$$D_{\mathbf{w}}(\mathbf{x}) = p_{\mathbf{w}}(y = 1|\mathbf{x}) = \sigma(F(\mathbf{w}, \mathbf{x})) = \frac{1}{1 + e^{-\mathbf{w}^\top \mathbf{x}}}$$

GANs with Linear Discriminator

- Linear scoring function $F(\mathbf{w}, \mathbf{x}) = \mathbf{w}^\top \mathbf{x}$, discriminator

$$D_{\mathbf{w}}(\mathbf{x}) = p_{\mathbf{w}}(y = 1|\mathbf{x}) = \sigma(F(\mathbf{w}, \mathbf{x})) = \frac{1}{1 + e^{-\mathbf{w}^\top \mathbf{x}}}$$

- Discriminator optimization problem:

$$\min_{\mathbf{w}} f(\theta, \mathbf{w}) = \min_{\mathbf{w}} \frac{C}{2} \|\mathbf{w}\|_2^2 + \frac{1}{2n} \sum_{i=1}^{n} \log\left(1 + e^{-\mathbf{w}^\top \mathbf{x}_i}\right) + \frac{1}{2n} \sum_{i=1}^{n} \log\left(1 + e^{\mathbf{w}^\top G_\theta(\mathbf{z}_i)}\right).$$

GANs with Linear Discriminator

- Linear scoring function $F(\mathbf{w}, \mathbf{x}) = \mathbf{w}^\top \mathbf{x}$, discriminator

$$D_{\mathbf{w}}(\mathbf{x}) = p_{\mathbf{w}}(y = 1|\mathbf{x}) = \sigma(F(\mathbf{w}, \mathbf{x})) = \frac{1}{1 + e^{-\mathbf{w}^\top \mathbf{x}}}$$

- Discriminator optimization problem:

real images

$$\min_{\mathbf{w}} f(\theta, \mathbf{w}) = \min_{\mathbf{w}} \frac{C}{2} \|\mathbf{w}\|_2^2 + \frac{1}{2n} \sum_{i=1}^{n} \log \left(1 + e^{-\mathbf{w}^\top \mathbf{x}_i}\right)$$

generated images

$$+ \frac{1}{2n} \sum_{i=1}^{n} \log \left(1 + e^{\mathbf{w}^\top G_\theta(\mathbf{z}_i)}\right).$$

GANs with Linear Discriminator

- Loss is convex in $\mathbf{w}$, the dual problem:

$$
\begin{aligned}
\max_{\lambda} \quad & g(\theta, \lambda) = -\frac{1}{2C}\left\|\sum_{i=1}^{n} \lambda_{\mathbf{x}_i}\mathbf{x}_i - \sum_{i=1}^{n} \lambda_{\mathbf{z}_i} G_\theta(\mathbf{z}_i)\right\|_2^2 \\
& \qquad + \frac{1}{2n}\sum_{i=1}^{n} H(2n\lambda_{\mathbf{x}_i}) + \frac{1}{2n}\sum_{i=1}^{n} H(2n\lambda_{\mathbf{z}_i}), \\
\text{s.t.} \quad & \forall i, \qquad 0 \le \lambda_{\mathbf{x}_i} \le \frac{1}{2n}, \quad 0 \le \lambda_{\mathbf{z}_i} \le \frac{1}{2n}.
\end{aligned}
$$

GANs with Linear Discriminator

- Loss is convex in $\mathbf{w}$, the dual problem:

weighted moment matching

$$\begin{aligned}
\max_{\lambda} \quad g(\theta, \lambda) = & -\frac{1}{2C} \left\| \sum_{i=1}^{n} \lambda_{\mathbf{x}_i} \mathbf{x}_i - \sum_{i=1}^{n} \lambda_{\mathbf{z}_i} G_\theta(\mathbf{z}_i) \right\|_2^2 \\
& + \frac{1}{2n} \sum_{i=1}^{n} H(2n\lambda_{\mathbf{x}_i}) + \frac{1}{2n} \sum_{i=1}^{n} H(2n\lambda_{\mathbf{z}_i}), \\
\text{s.t.} \quad \forall i, \quad & 0 \leq \lambda_{\mathbf{x}_i} \leq \frac{1}{2n}, \quad 0 \leq \lambda_{\mathbf{z}_i} \leq \frac{1}{2n}.
\end{aligned}$$

GANs with Linear Discriminator

- Loss is convex in $\mathbf{w}$, the dual problem:

weighted moment matching

$$\begin{aligned}
\max_{\lambda} \quad & g(\theta, \lambda) = -\frac{1}{2C}\left\| \sum_{i=1}^{n} \lambda_{\mathbf{x}_i}\mathbf{x}_i - \sum_{i=1}^{n} \lambda_{\mathbf{z}_i} G_\theta(\mathbf{z}_i) \right\|_2^2 \\
& \qquad + \frac{1}{2n}\sum_{i=1}^{n} H(2n\lambda_{\mathbf{x}_i}) + \frac{1}{2n}\sum_{i=1}^{n} H(2n\lambda_{\mathbf{z}_i}), \\
\text{s.t.} \quad & \forall i, \quad 0 \le \lambda_{\mathbf{x}_i} \le \frac{1}{2n}, \quad 0 \le \lambda_{\mathbf{z}_i} \le \frac{1}{2n}.
\end{aligned}$$

entropy

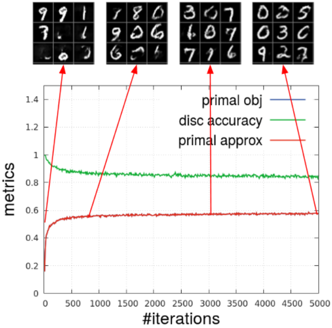
GANs with Linear Discriminator

- Alternate $\theta \leftarrow \theta + \eta_\theta \nabla_\theta g(\theta, \lambda)$ and $\lambda = \mathrm{argmax}_{\lambda'}\, g(\theta, \lambda)$.

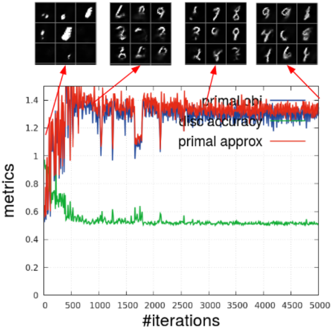
- Solving the dual is not hard (quadratic optimization).

GANs with Linear Discriminator

- Alternate $\theta \leftarrow \theta + \eta_\theta \nabla_\theta g(\theta, \lambda)$ and $\lambda = \text{argmax}_{\lambda'}\, g(\theta, \lambda)$.
- Solving the dual is not hard (quadratic optimization).
- Learning is very stable:



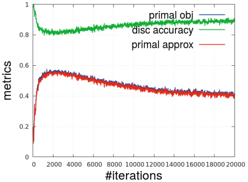
Dual GAN



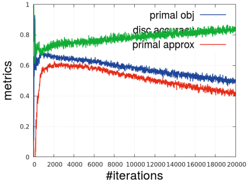
Standard GAN

GANs with Non-Linear Discriminator

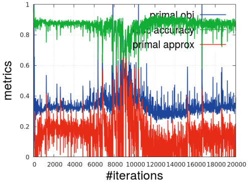
Improved stability and sensitivity to hyperparameters.



Dual GAN Score Lin.



Dual GAN Cost Lin.



Standard GAN

Dualing GANs

Today Pacific Ballroom #103

See you at our poster.