Dualing GANs

Yujia Li\textsuperscript{1,*}, Alexander Schwing\textsuperscript{3}, Kuan-Chieh Wang\textsuperscript{1,2} and Richard Zemel\textsuperscript{1,2}

\textsuperscript{1}University of Toronto \hspace{1cm} \textsuperscript{2}Vector Institute
\textsuperscript{3}University of Illinois at Urbana-Champaign \hspace{1cm} *Now at DeepMind
Instability in GAN Training

- Saddle point formulation for GAN training:

\[
\max_{\theta} \min_{w} f(\theta, w)
\]
Instability in GAN Training

- Saddle point formulation for GAN training:
  \[
  \max_{\theta} \min_{w} f(\theta, w)
  \]

- Alternate gradient updates for \(\theta\) and \(w\)
  \[
  \theta \rightarrow \theta + \eta_{\theta} \nabla_{\theta} f(\theta, w), \quad w \rightarrow w - \eta_{w} \nabla_{w} f(\theta, w).
  \]

- This leads to instability.
Proposed solution: $\max_\theta \min_w \rightarrow \max_\theta \max_\lambda$

- Dualize $\min_w f(\theta, w)$ into $\max_\lambda g(\theta, \lambda)$.

- $\max_\theta \max_\lambda g(\theta, \lambda)$ is more stable than $\max_\theta \min_w f(\theta, w)$. 
GANs with Linear Discriminator

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

\[
D_w(x) = p_w(y = 1|z) = \sigma(F(w, x)) = \frac{1}{1 + e^{-w^\top x}}
\]
GANs with Linear Discriminator

- Linear scoring function \( F(w, x) = w^Tx \), discriminator

\[
D_w(x) = p_w(y = 1|x) = \sigma(F(w, x)) = \frac{1}{1 + e^{-w^Tx}}
\]

- Discriminator optimization problem:

\[
\min_w f(\theta, w) = \min_w \frac{C}{2} \|w\|_2^2 + \frac{1}{2n} \sum_{i=1}^n \log \left(1 + e^{-w^T x_i}\right) + \frac{1}{2n} \sum_{i=1}^n \log \left(1 + e^{w^T G_\theta(z_i)}\right).
\]
GANs with Linear Discriminator

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

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

- Discriminator optimization problem:

$$\min_w f(\theta, w) = \min_w \frac{C}{2} \|w\|^2 + \frac{1}{2n} \sum_{i=1}^{n} \log \left(1 + e^{-w^\top x_i}\right)$$

real images

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

generated images
GANs with Linear Discriminator

- Loss is convex in $w$, the dual problem:

$$
\max_{\lambda} \quad g(\theta, \lambda) = -\frac{1}{2C} \left\| \sum_{i=1}^{n} \lambda_{x_i} x_i - \sum_{i=1}^{n} \lambda_{z_i} G_{\theta}(z_i) \right\|_2^2
$$

$$
+ \frac{1}{2n} \sum_{i=1}^{n} H(2n\lambda_{x_i}) + \frac{1}{2n} \sum_{i=1}^{n} H(2n\lambda_{z_i}),
$$

s.t. \quad \forall i, \quad 0 \leq \lambda_{x_i} \leq \frac{1}{2n}, \quad 0 \leq \lambda_{z_i} \leq \frac{1}{2n}.$$
GANs with Linear Discriminator

- Loss is convex in $w$, the dual problem:

$$
\max_{\lambda} \quad g(\theta, \lambda) = -\frac{1}{2C} \left\| \sum_{i=1}^{n} \lambda_{x_i} x_i - \sum_{i=1}^{n} \lambda_{z_i} G_\theta(z_i) \right\|_2^2 \\
+ \frac{1}{2n} \sum_{i=1}^{n} H(2n\lambda_{x_i}) + \frac{1}{2n} \sum_{i=1}^{n} H(2n\lambda_{z_i}),
$$

s.t. \quad \forall i, \quad 0 \leq \lambda_{x_i} \leq \frac{1}{2n}, \quad 0 \leq \lambda_{z_i} \leq \frac{1}{2n}.

weighted moment matching
GANs with Linear Discriminator

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

$$
\max_{\lambda} \quad g(\theta, \lambda) = -\frac{1}{2C} \left\| \sum_{i=1}^{n} \lambda_{x_i} x_i - \sum_{i=1}^{n} \lambda_{z_i} G_{\theta}(z_i) \right\|^2_2 \\
+ \frac{1}{2n} \sum_{i=1}^{n} H(2n\lambda_{x_i}) + \frac{1}{2n} \sum_{i=1}^{n} H(2n\lambda_{z_i})
$$

s.t. \quad \forall i, \quad 0 \leq \lambda_{x_i} \leq \frac{1}{2n}, \quad 0 \leq \lambda_{z_i} \leq \frac{1}{2n}.

weighted moment matching

entropy
GANs with Linear Discriminator

- Alternate $\theta \leftarrow \theta + \eta_\theta \nabla_\theta g(\theta, \lambda)$ and $\lambda = \arg\max_\lambda g(\theta, \lambda)$.

- Solving the dual is not hard (quadratic optimization).
GANs with Linear Discriminator

- Alternate $\theta \leftarrow \theta + \eta \nabla_\theta g(\theta, \lambda)$ and $\lambda = \arg\max_\lambda g(\theta, \lambda)$.

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

- Learning is very stable:
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.