# Slide material – Section 4 – John Flynn

*Firstly, hopefully someone else has this equation in their slides before mine!*

$$\min_G \max_D V(D, G) = \mathbb{E}_{\boldsymbol{x} \sim p_{\text{data}}(\boldsymbol{x})}[\log D(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{z} \sim p_{\boldsymbol{z}}(\boldsymbol{z})}[\log(1 - D(G(\boldsymbol{z})))]. \tag{1}$$

Figure 1: Equation 1

*And Algorithm 1 pseudo code description from page 4.*

# Global optimum for $p_g = p_{data}$

**Prove**: For G fixed the optimal discriminator D is
$D*_G(x) = \frac{p_{\text{data}}(x)}{p_{\text{data}}(x)+p_{\text{gen}}(x)}$

For any non-zero $(a, b) \in R^2$, the function
$y \to a\log(y) + b\log(1-y)$ achieves its maximum in $[0, 1]$ at $\frac{a}{a+b}$.

- Demonstration:
  https://www.desmos.com/calculator/odj1obavmt

**Theorem 1 summary**: Global minimum achieved $\iff p_g = p_{data}$

- Since $\frac{a}{a+a} = 1/2$, if $p_g = p_{data}$ then $D*_G(x) = 1/2$.
- Best value for $V(D*_G(x)) = -\log 4$.
- View Kullback-Leibler divergences between $p_g$ and $p_{data}$ as Jenson-Shannon divergence.
- JSD zero only when equal, so remaining term must be $-\log 4$.
- If $p_g \neq p_{data}$, JSD will be non-zero positive, and will add to the $-\log 4$, i.e. **not the minimum**.

# Convergence of algorithm

**Proposition 2**: If G and D have enough capacity, and at each step of Algorithm 1, D reaches optimum given G, and $p_g$ is updated to improve the criterion then $p_g$ converges to $p_{data}$

- Important assumption: G and D need enough capacity!
- Authors prove by showing:
- $V(G, D)$ is **convex** as distribution $p_g$ is varied.
  - Theorem 1 showed maximum value is unique, so all other values must be less i.e. $V(G, D)$ is convex.
- If it's convex, a small enough learning rate will update $p_g$ allowing it to converge to $p_{data}$