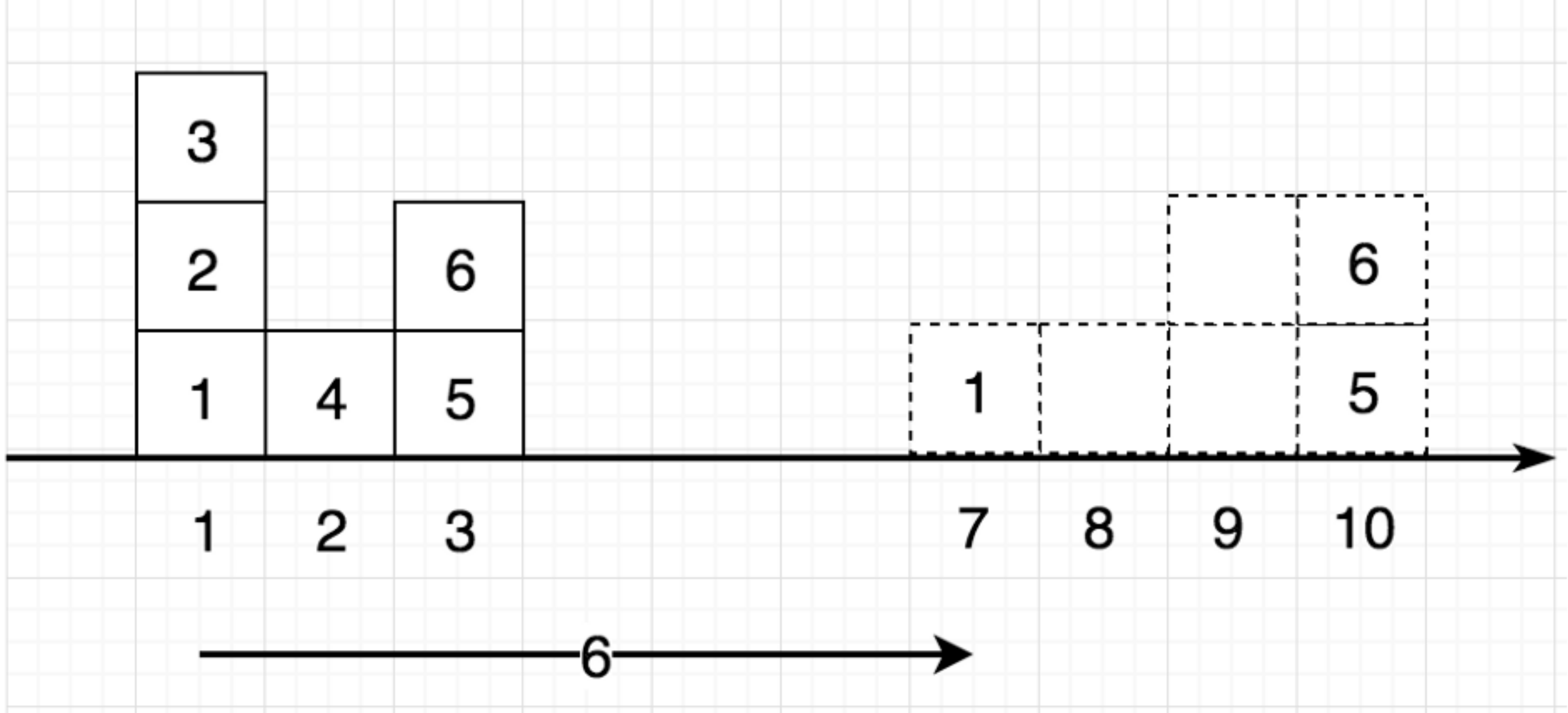
GANs special edition

WassersteinGAN

Mikhail Usvyatsov
20.10.2018

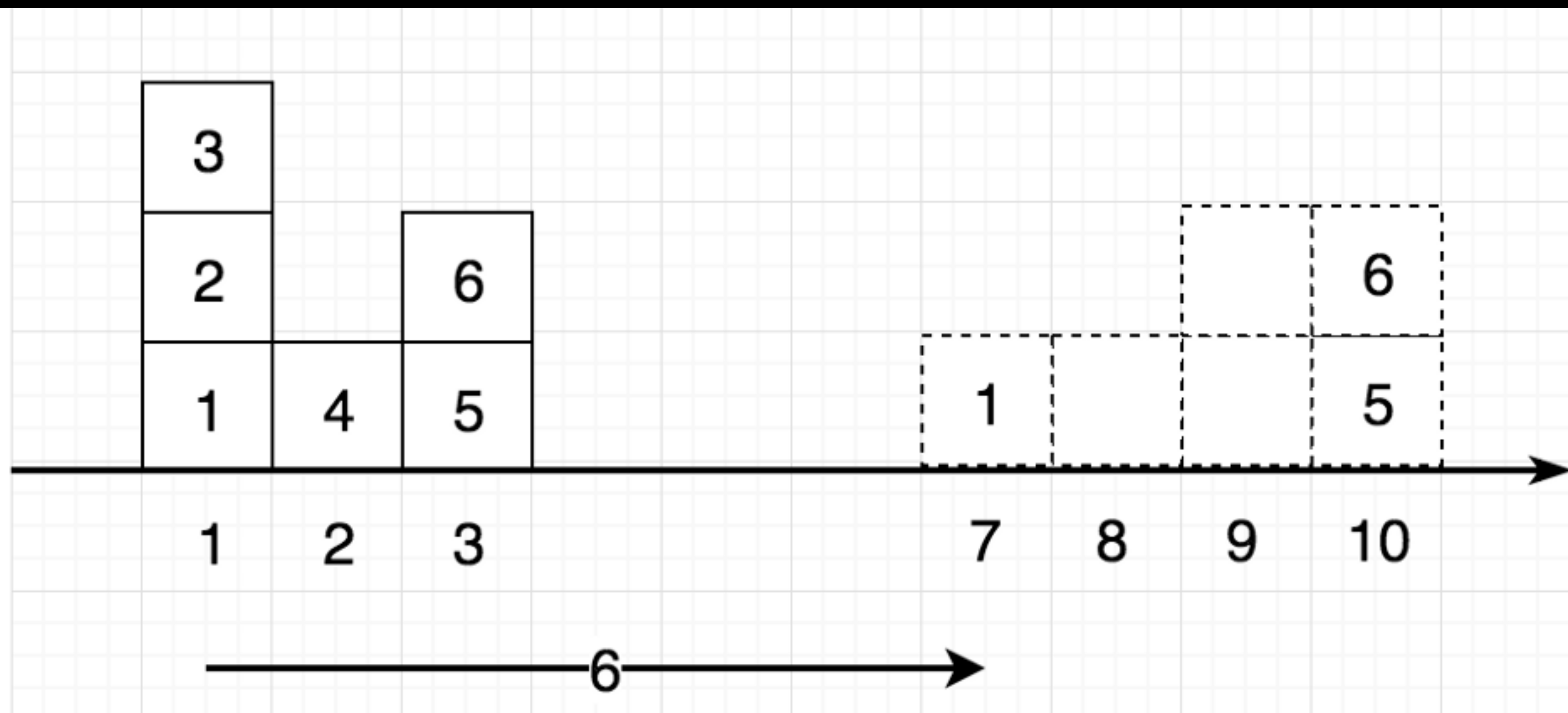
Earth-Mover distance (EMD) / Wasserstein Metric



The moving cost equals to its weight times the distance. For simplicity, we will set the weight to be 1.

Quiz: What is the cost to move box #1 to a new position ?

Earth-Mover distance (EMD) / Wasserstein Metric



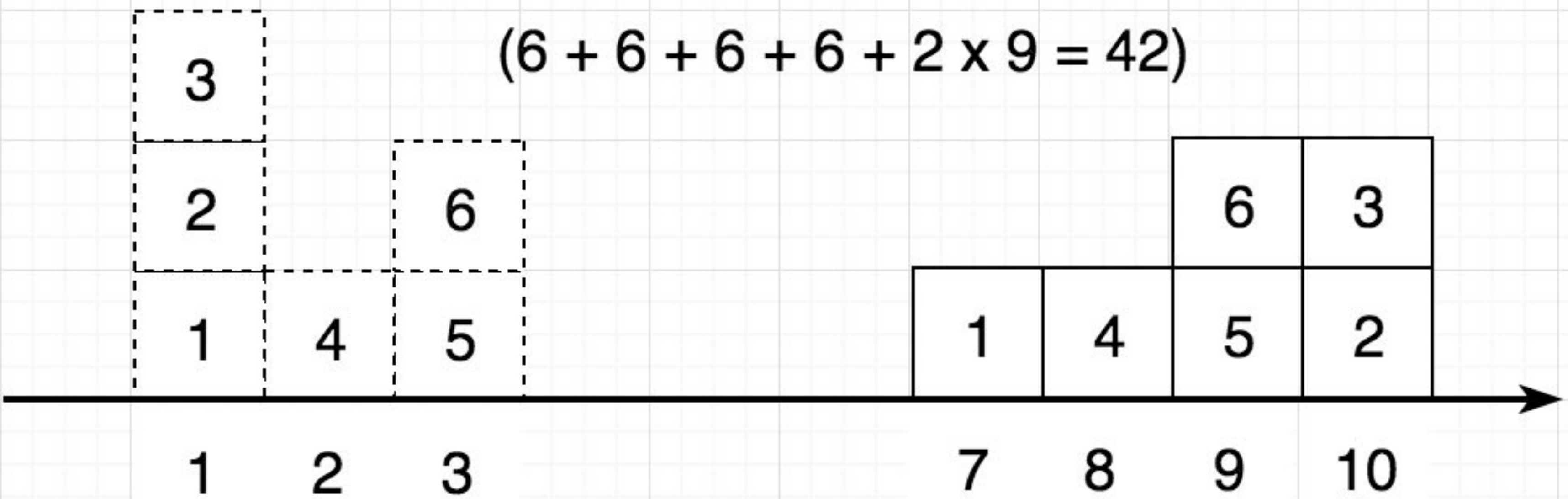
The moving cost equals to its weight times the distance. For simplicity, we will set the weight to be 1.

Quiz: What is the cost to move box #1 to a new position ?

Answer: 6 (7–1)

Earth-Mover distance (EMD) / Wasserstein Metric

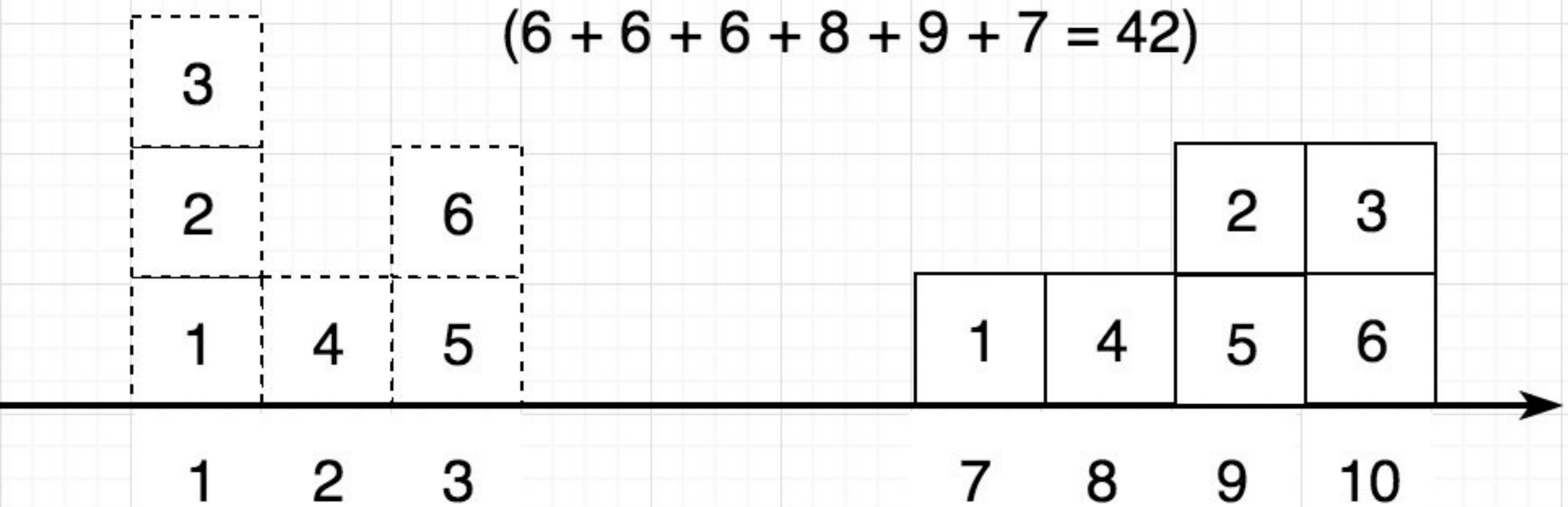
$(6 + 6 + 6 + 6 + 2 \times 9 = 42)$



γ_1

	7	8	9	10
1	1	0	0	2
2	0	1	0	0
3	0	0	2	0

$(6 + 6 + 6 + 8 + 9 + 7 = 42)$



γ_2

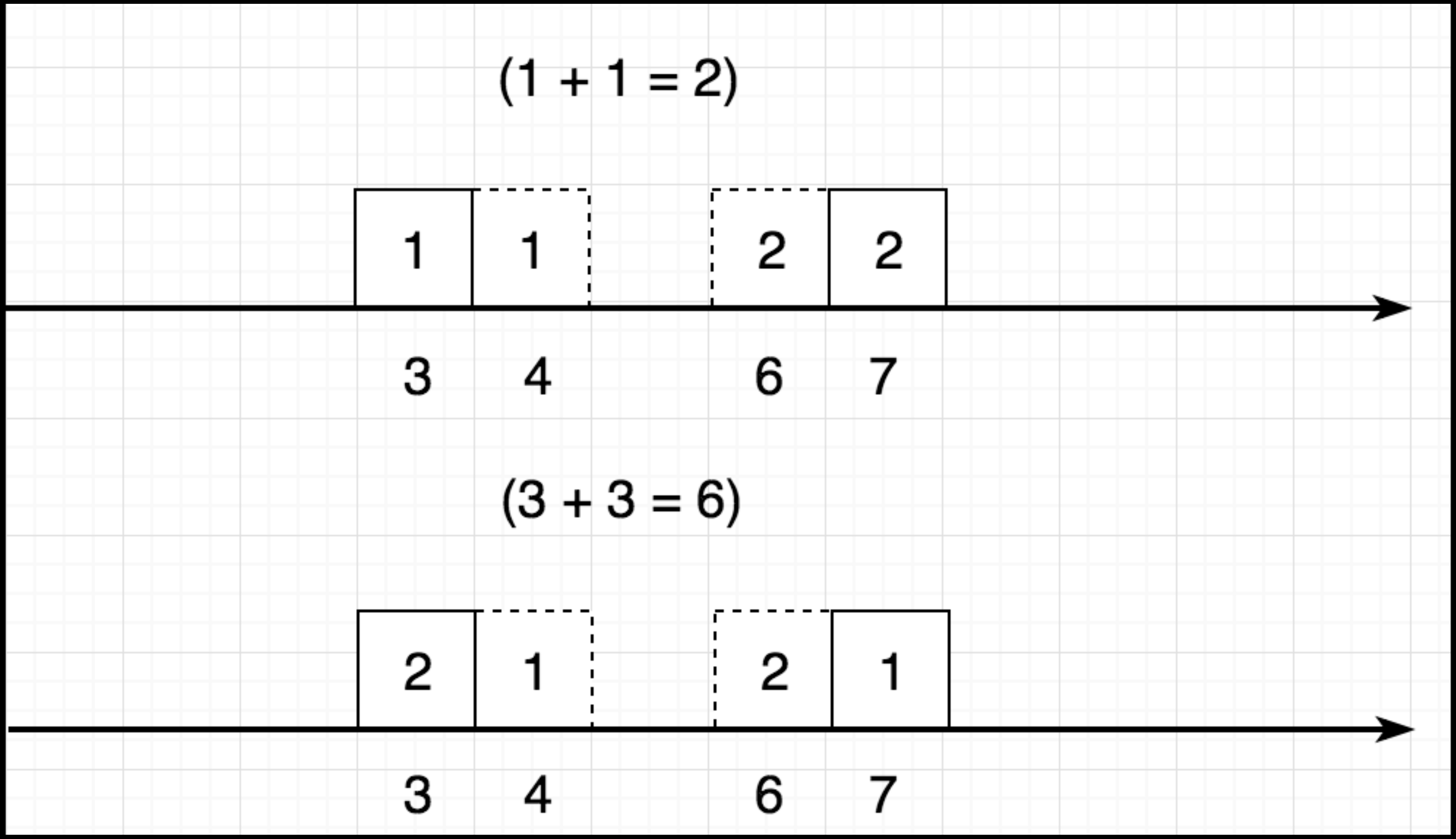
	7	8	9	10
1	1	0	1	1
2	0	0	1	0
3	0	0	1	1

Earth-Mover distance (EMD) / Wasserstein Metric

The **Wasserstein distance** (or the **EMD**) is the cost of the cheapest transport plan.

Earth-Mover distance (EMD) / Wasserstein Metric

The **Wasserstein distance** (or the **EMD**) is the cost of the cheapest transport plan.



Earth-Mover distance (EMD) / Wasserstein Metric

The Wasserstein distance is the minimum cost of transporting mass in converting the data distribution ***q*** to the data distribution ***p***.

The Wasserstein distance for the real data distribution ***Pr*** and the generated data distribution ***Pg*** is mathematically defined as the greatest lower bound (infimum) for any transport plan (i.e. the cost for the cheapest plan):

Earth-Mover distance (EMD) / Wasserstein Metric

The Wasserstein distance is the minimum cost of transporting mass in converting the data distribution ***q*** to the data distribution ***p***.

The Wasserstein distance for the real data distribution ***Pr*** and the generated data distribution ***Pg*** is mathematically defined as the greatest lower bound (infimum) for any transport plan (i.e. the cost for the cheapest plan):

$$W(P_r, P_g) = \inf_{\gamma \in \Pi(P_r, P_g)} E_{(x,y)\sim\gamma}[||x - y||]$$

Π(Pr, Pg) denotes the set of all joint distributions γ(x, y) whose marginals are respectively Pr and Pg.

Earth-Mover distance (EMD) / Wasserstein Metric

Π contains all the possible transport plan ***γ***.

y

Π

x		7	8	9	10
	1	1	0	0	2
x	2	0	1	0	0
	3	0	0	2	0

γ_1

x		7	8	9	10
	1	1	0	0	2
x	2	0	0	1	0
	3	0	1	1	0

γ_2

Earth-Mover distance (EMD) / Wasserstein Metric

We combine variable *x* and *y* to form a joint distribution *γ(x, y)* and *γ(1, 10)* is simply how many boxes at location 10 is from location 1. The number of boxes in location 10 must originally come from any position, i.e. *∑ γ(*, 10) = 2.* That is the same as saying *γ(x, y)* must have marginals ***Pr*** and ***Pg*** respectively.

y Π

x \ y	7	8	9	10
1	1	0	0	2
2	0	1	0	0
3	0	0	2	0

γ_1

x \ y	7	8	9	10
1	1	0	0	2
2	0	0	1	0
3	0	1	1	0

γ_2

Arjovsky’ way

Arjovsky et al 2017 wrote a paper to illustrate the GAN problem mathematically with the following conclusions:

> An optimal discriminator produces good information for the generator to improve.
> But if the generator is not doing a good job yet, the gradient for the generator diminishes and the generator learns nothing

Arjovsky’ way

- Solution 1 (proposed by Godfellow in the original paper)

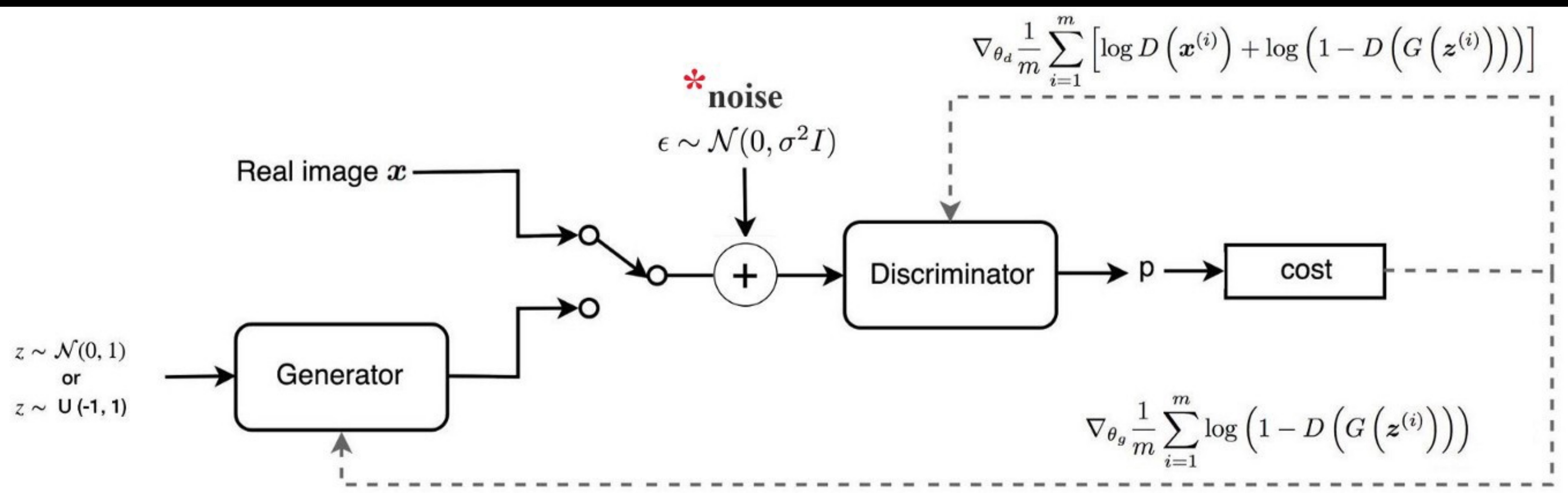
$$L = -\,log(D(G(Z)))$$

Arjovsky’ way

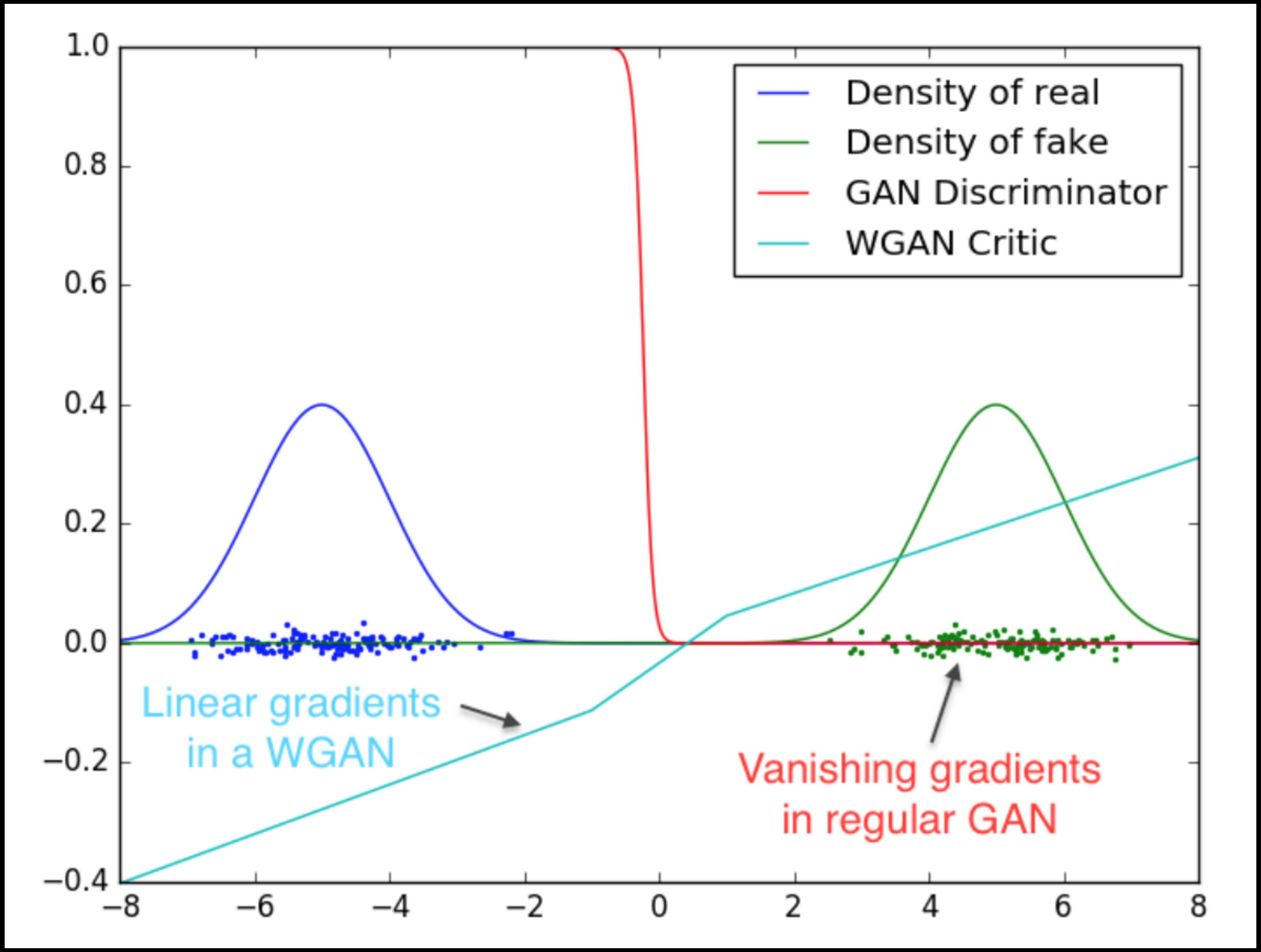
- Solution 1 (proposed by Godfellow in the original paper)

$L = -\log(D(G(Z)))$

- Solution 2



Wasserstein GAN



Wasserstein GAN

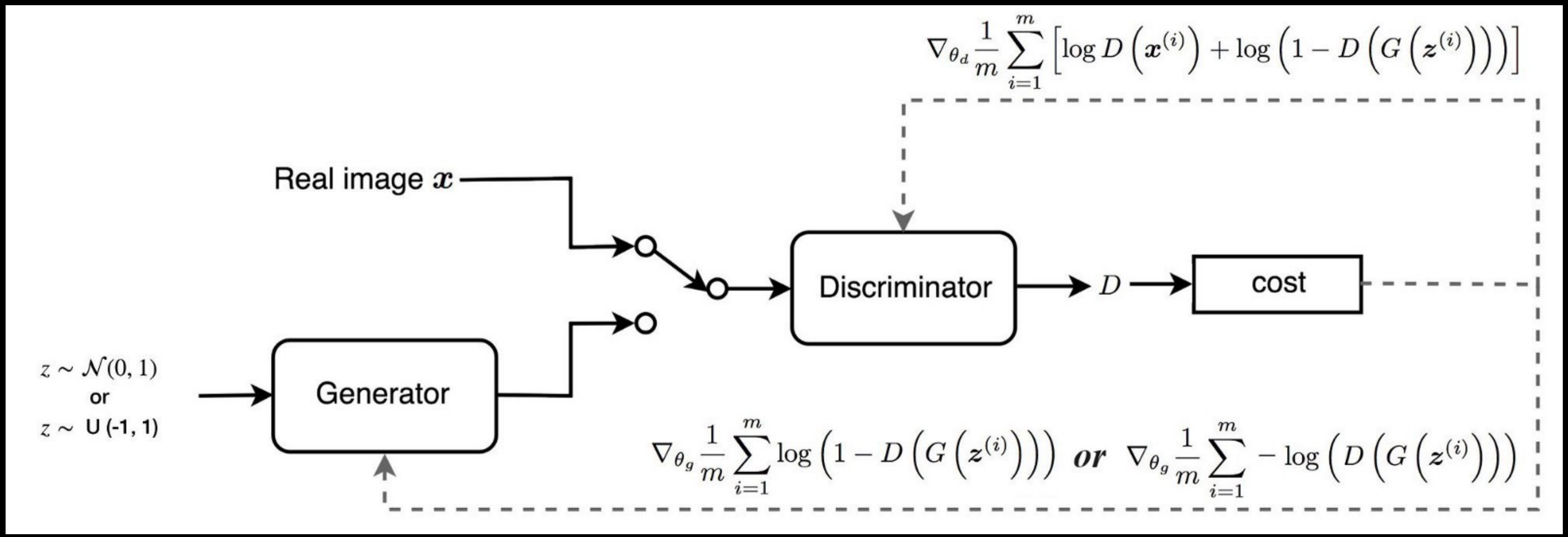
However, the equation for the Wasserstein distance is highly intractable. Using the Kantorovich-Rubinstein duality, we can simplify the calculation to

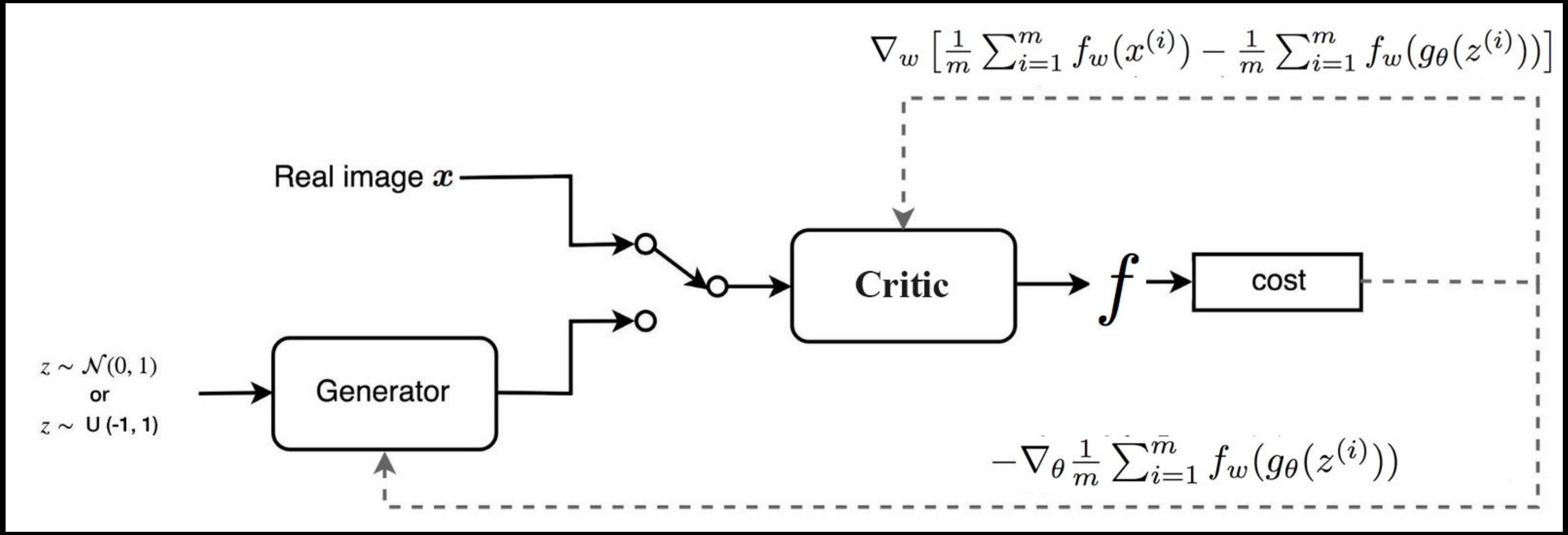
$$W(P_r, P_g) = \sup_{||f||_l \leq 1} E_{x \sim P_r}[f(x)] - E_{x \sim P_\theta}[f(x)]$$

where *sup* is the least upper bound and ***f*** is a 1-Lipschitz function following this constraint:

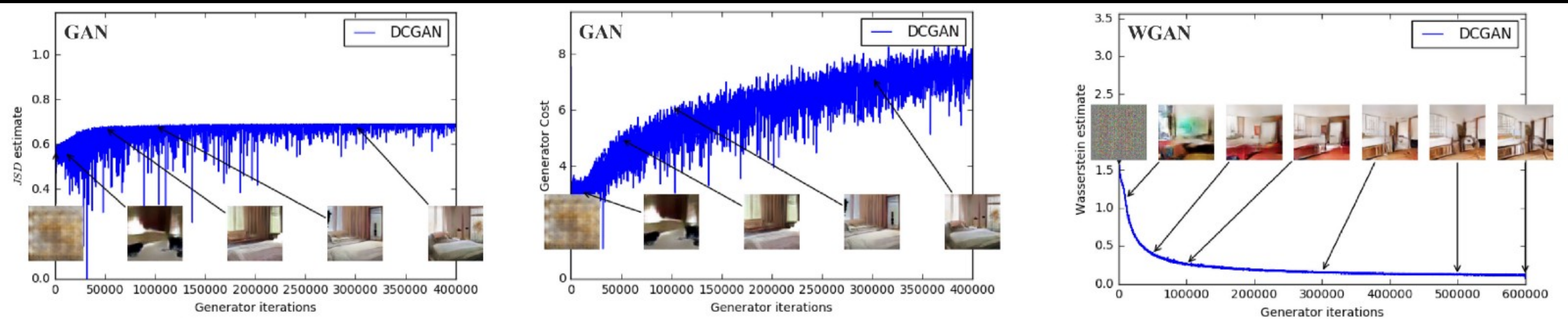
$$|f(x_1) - f(x_2)| \leq |x_1 - x_2|$$

Wasserstein GAN





Wasserstein GAN



$$\frac{1}{m}\sum_{i=1}^{m}\log\left(1-D\left(G\left(z^{(i)}\right)\right)\right)$$

$$\frac{1}{m}\sum_{i=1}^{m}-\log\left(D\left(G\left(z^{(i)}\right)\right)\right)$$

$$\frac{1}{m}\sum_{i=1}^{m} f_w(x^{(i)}) - \frac{1}{m}\sum_{i=1}^{m} f_w(g_\theta(z^{(i)}))$$

Wasserstein GAN With gradient penalty

$$L = \underbrace{\underset{\tilde{\boldsymbol{x}} \sim \mathbb{P}_g}{\mathbb{E}} \left[D(\tilde{\boldsymbol{x}}) \right] - \underset{\boldsymbol{x} \sim \mathbb{P}_r}{\mathbb{E}} \left[D(\boldsymbol{x}) \right]}_{\text{Original critic loss}} + \underbrace{\lambda \underset{\hat{\boldsymbol{x}} \sim \mathbb{P}_{\hat{\boldsymbol{x}}}}{\mathbb{E}} \left[(\| \nabla_{\hat{\boldsymbol{x}}} D(\hat{\boldsymbol{x}}) \|_2 - 1)^2 \right]}_{\text{Our gradient penalty}} .$$

where $\hat{\boldsymbol{x}}$ ***sampled from*** $\tilde{\boldsymbol{x}}$ ***and*** $\boldsymbol{x}$ ***with t uniformly sampled between 0 and 1***

$$\hat{\boldsymbol{x}} = t\tilde{\boldsymbol{x}} + (1-t)\boldsymbol{x} \text{ with } 0 \leq t \leq 1$$

Wasserstein GAN With gradient penalty