Surfing: Iterative Optimization Over Incrementally Trained Deep Networks

Ganlin Song, Zhou Fan, John Lafferty

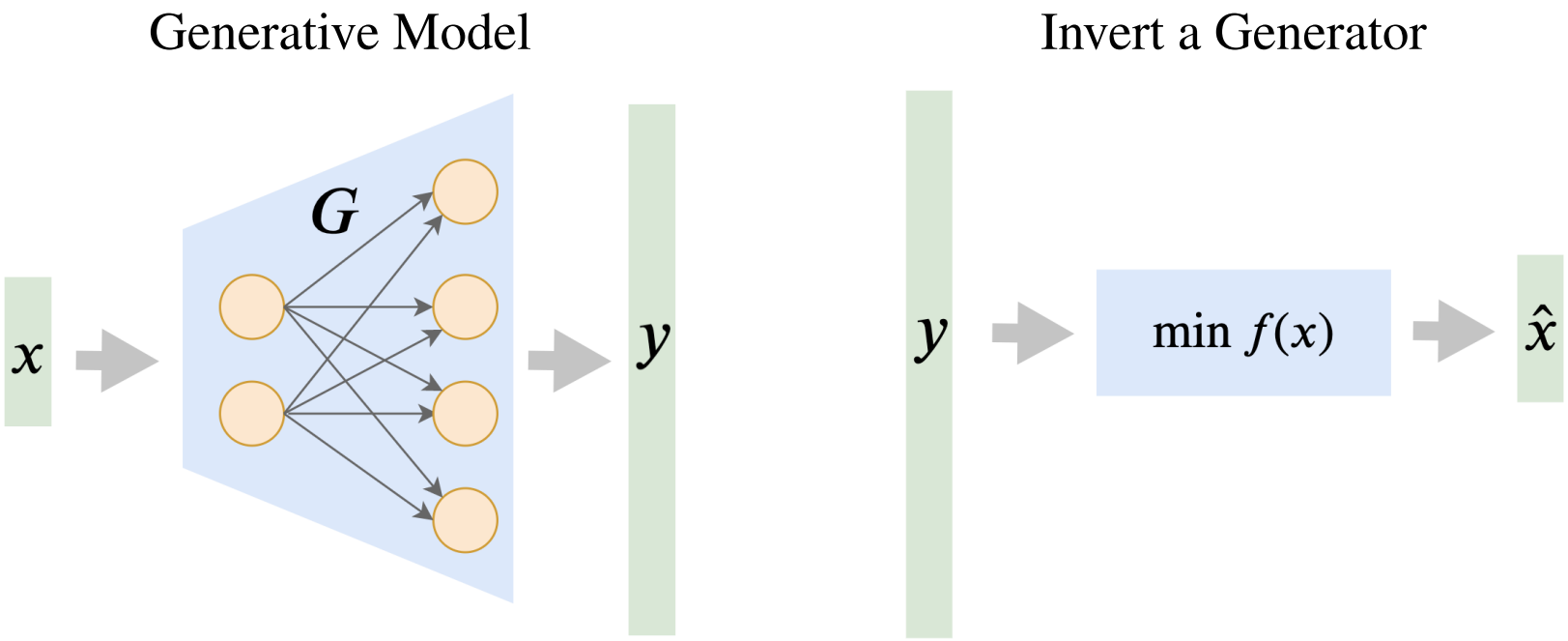
Department of Statistics and Data Science

Yale University

Background

We consider inverting a trained generative network G by

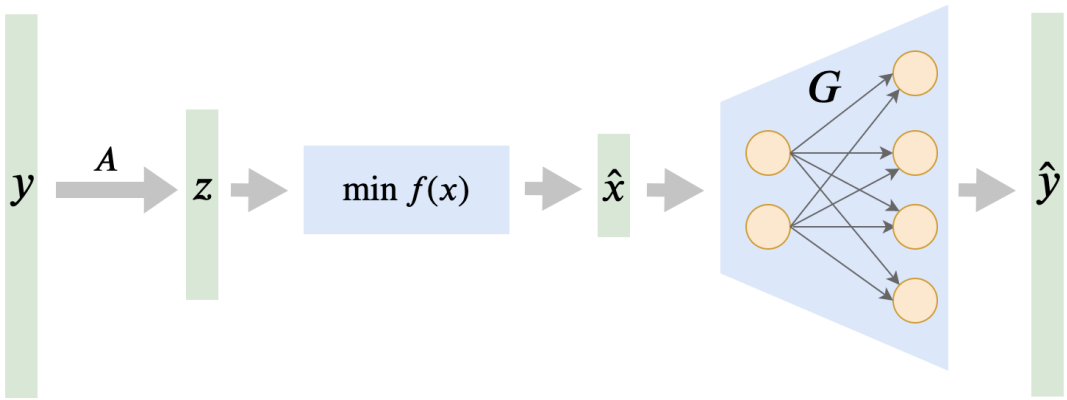
$$\min_x f(x) = \min_x \|G(x) - y\|^2$$



Background

- Compressed sensing framework: observe $z = Ay + \epsilon$; recover y by
  (Bora, Jalal, Price & Dimakis 2017)

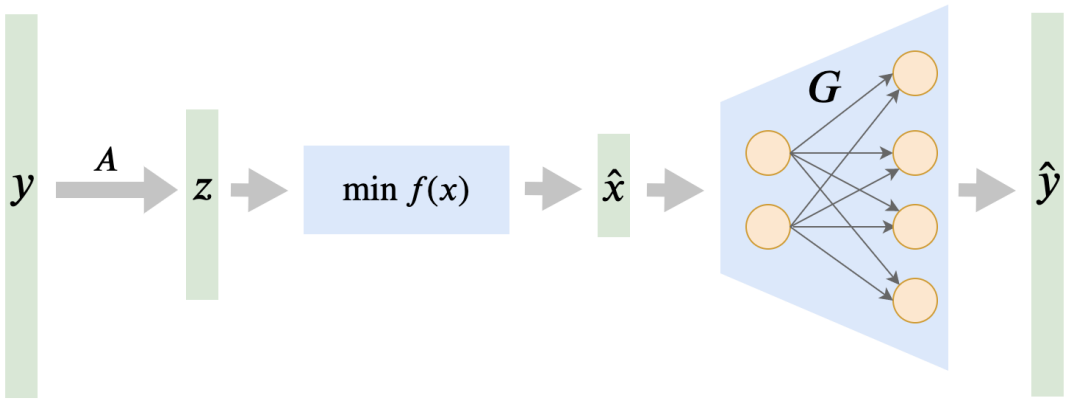
$$\min_x f(x) = \min_x \|AG(x) - z\|^2$$



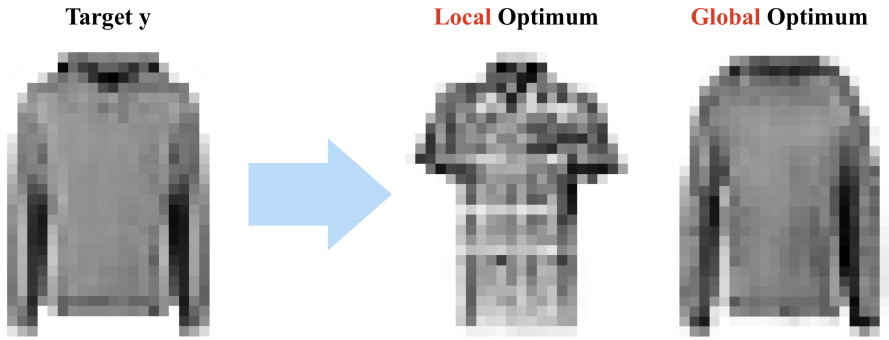
Background

- Compressed sensing framework: observe $z = Ay + \epsilon$; recover y by
  (Bora, Jalal, Price & Dimakis 2017)

$$\min_x f(x) = \min_x \|AG(x) - z\|^2$$

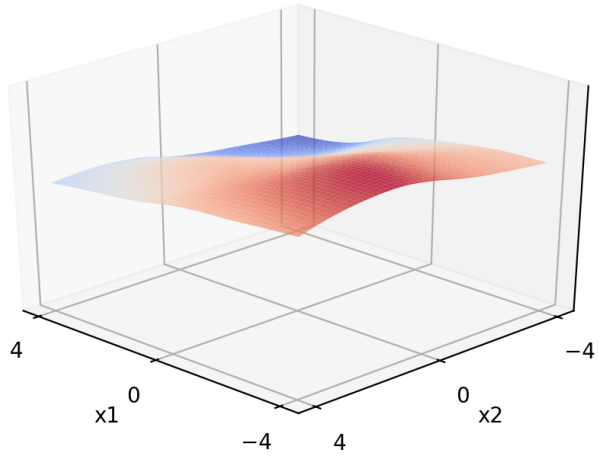


- $f(x)$ is non-convex; gradient descent not guaranteed to reach global optimum

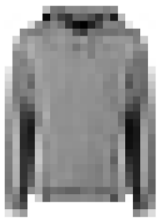


Motivation

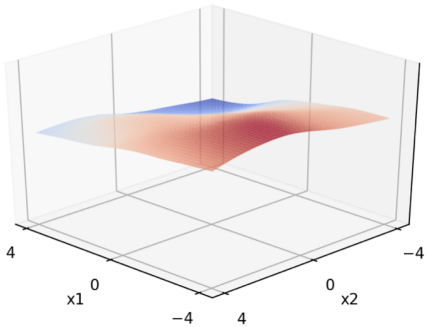
Landscape of $x \mapsto -f_\theta(x) = -\|G_\theta(x) - y\|^2$, as weights θ are trained



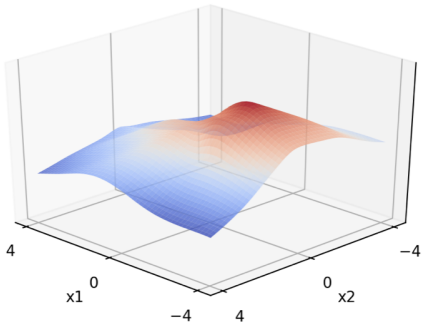
Target y



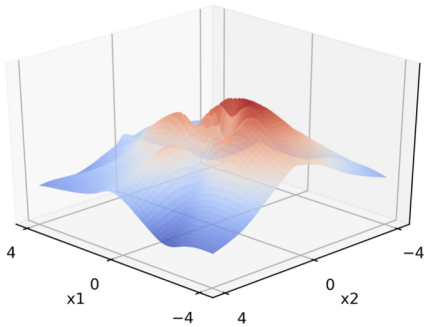
Initial Network



Partially Trained



Fully Trained



Algorithm

Intuition

- The landscape for initial random network is "nice"
- Initialize with random network and track optimum for intermediate networks

Surfing Algorithm

- Obtain a sequence of parameters $\theta_0, \theta_1, \ldots, \theta_T$ during training
- Optimize empirical risk function $f_{\theta_0}, f_{\theta_1}, \ldots, f_{\theta_T}$ iteratively using gradient descent
- For each $t \in \{1, \ldots, T\}$, initialize gradient descent at the solution from time $t-1$

Theory and Experiments

Theoretical Results

1. If G_θ has random parameters, all critical points of $f_\theta(x)$ belong to a small neighborhood around 0 with high probability (Builds on Hand & Voroninski 2017)
2. Under certain conditions, modified surfing can track the minimizer

Theory and Experiments

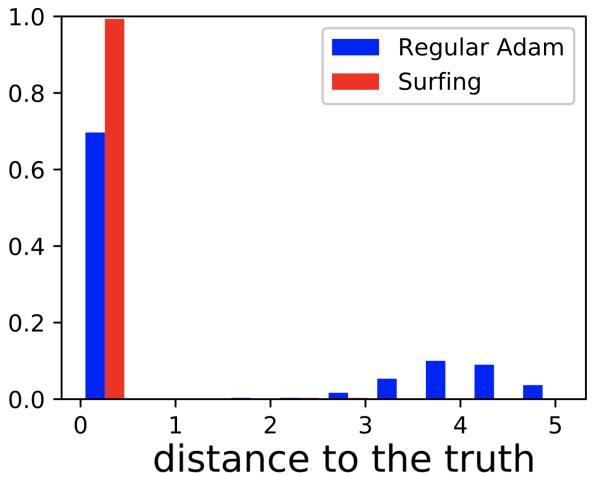
Theoretical Results

1. If G_θ has random parameters, all critical points of $f_\theta(x)$ belong to a small neighborhood around 0 with high probability (Builds on Hand & Voroninski 2017)
2. Under certain conditions, modified surfing can track the minimizer

Experiments

For DCGAN trained on Fashion-MNIST

$\min_x \|G_\theta(x) - G_\theta(x_0)\|^2$



$\min_x \|AG_\theta(x) - Ay\|^2$