Stochasticity of Deterministic Gradient Descent: Quantitative Local Min Escape in Multiscale Landscape

Molei Tao

School of Mathematics, Georgia Tech, USA

November 12, 2023 Caltech ACM

Machine Learning & Applied Computational Math

Machine Learning & Applied Computational Math

→ ML for problems in computing, sciences & engineering

Machine Learning & Applied Computational Math

$\rightarrow$ ML for problems in computing, sciences & engineering

$\leftarrow$ how ACM helps design and analyze
optimization, sampling, and deep learning practices

Machine Learning & Applied Computational Math

→ ML for problems in computing, sciences & engineering

← how ACM helps design and analyze optimization, sampling, and deep learning practices

*common belief: **large learning rate** is good*

Machine Learning & Applied Computational Math

$\rightarrow$ ML for problems in computing, sciences & engineering

$\leftarrow$ how ACM helps design and analyze optimization, sampling, and deep learning practices

common belief: ***large learning rate*** *is good*

$\min f(x)$ $$x_{k+1} = x_k - \boxed{h} \nabla f(x_k)$$ Gradient Descent

Machine Learning & Applied Computational Math

$\rightarrow$ ML for problems in computing, sciences & engineering

$\leftarrow$ how ACM helps design and analyze optimization, sampling, and deep learning practices

common belief: ***large learning rate*** *is good*

1st thought:

$\min f(x)$

$$x_{k+1} = x_k - h\nabla f(x_k)$$ Gradient Descent

$$\Updownarrow \quad x_i \approx x(ih)$$

$$\dot{x} = -\nabla f(x)$$ Gradient Flow

Machine Learning & Applied Computational Math

$\rightarrow$ ML for problems in computing, sciences & engineering

$\leftarrow$ how ACM helps design and analyze optimization, sampling, and deep learning practices

common belief: ***large learning rate*** *is good*

1st thought:

$\min f(x)$ $\quad x_{k+1} = x_k - h\nabla f(x_k)$ Gradient Descent

$\Updownarrow \quad x_i \approx x(ih)$

x(T) with fixed T

$\downarrow$

[T/h] steps

$\downarrow$

large h?

$\dot{x} = -\nabla f(x)$ Gradient Flow

Machine Learning & Applied Computational Math

→ ML for problems in computing, sciences & engineering

← how ACM helps design and analyze
optimization, sampling, and deep learning practices

*common belief: **large learning rate** is good*

1st thought:

$\min f(x)$

$$x_{k+1} = x_k - h\nabla f(x_k)$$

Gradient Descent

$$\Updownarrow \quad \cancel{x_i \approx x(ih)}$$

x(T) with fixed T
↓
[T/h] steps
↓
large h?

$$\dot{x} = -\nabla f(x)$$

Gradient Flow

Machine Learning & Applied Computational Math

$\rightarrow$ ML for problems in computing, sciences & engineering

$\leftarrow$ how ACM helps design and analyze
optimization, sampling, and deep learning practices

common belief: ***large learning rate*** *is good*

not just faster, but implicitly bias toward (desirable) global structures

???

Machine Learning & Applied Computational Math

$\rightarrow$ ML for problems in computing, sciences & engineering

$\leftarrow$ how ACM helps design and analyze
optimization, sampling, and deep learning practices

*common belief: **large learning rate** is good*

not just faster, but implicitly bias toward (desirable) global structures

???

→ better training & better test accuracies

training

model (e.g., neural network) parameters

$$\min_{x} f(x) := \sum_{i} d(\mathrm{output}_i, g(x, \mathrm{input}_i))$$

training data (known)

training

model (e.g., neural network) parameters

$$\min_{x} f(x) := \sum_i d(\mathrm{output}_i, g(x, \mathrm{input}_i))$$

training data (known)

trapped in local min <-> suboptimal training accuracy

training

model (e.g., neural network) parameters

$$\min_{x} f(x) := \sum_i d(\mathrm{output}_i, g(x, \mathrm{input}_i))$$

training data (known)

large h? this talk

trapped in local min ←→ suboptimal training accuracy

training

model (e.g., neural network) parameters

$$\min_{x} f(x) := \sum_{i} d(\text{output}_i, g(x, \text{input}_i))$$

training data (known)

large h? this talk

trapped in local min ←→ suboptimal training accuracy

testing

matters

$$d(\widehat{\text{output}}_j, g(x_{\text{trained}}, \widehat{\text{input}}_j))$$

test data

training

model (e.g., neural network) parameters

$$\min_{x} f(x) := \sum_i d(\text{output}_i, g(x, \text{input}_i))$$

training data (known)

large h? this talk

trapped in local min $\leftarrow\rightarrow$ suboptimal training accuracy

generalization

testing

matters

$$d(\widehat{\text{output}}_j, g(x_{\text{trained}}, \widehat{\text{input}}_j))$$

test data

model (e.g., neural network) parameters

training

$$\min_{x} f(x) := \sum_i d(\text{output}_i, g(x, \text{input}_i))$$

training data (known)

large h? this talk

trapped in local min ←→ suboptimal training accuracy

generalization

large h? offline

matters

testing

$$d(\widehat{\text{output}_j}, g(x_{\text{trained}}, \widehat{\text{input}_j}))$$

test data

Stochasticity of Deterministic Gradient Descent: Large Learning Rate for Multiscale Objective Function

NeurIPS 2020, arXiv: 2002.06189

Lingkai Kong[1] and *Molei Tao*[1]

1 Georgia Institute of Technology (USA)

$\min f(x)$

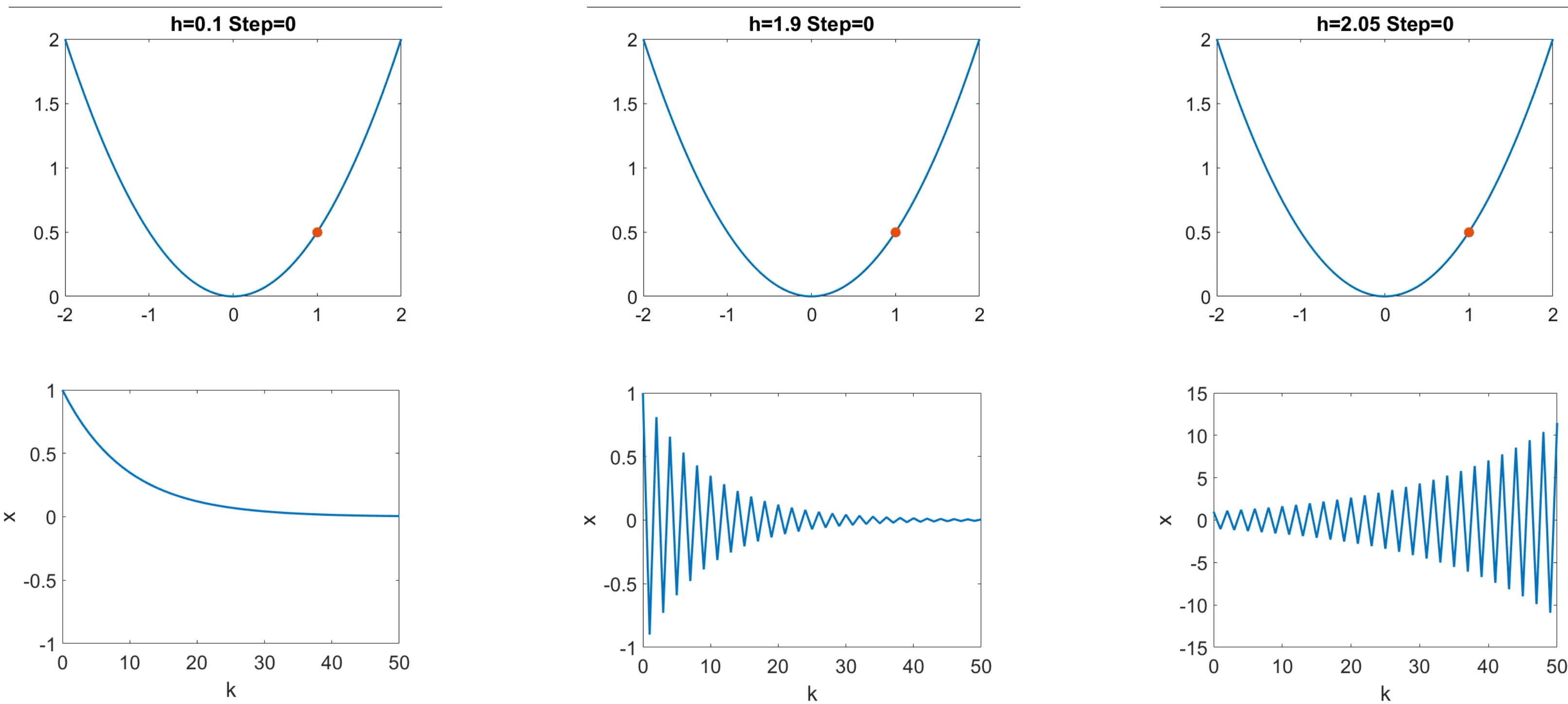
$$x_{k+1} = x_k - h\nabla f(x_k)$$

Gradient Descent

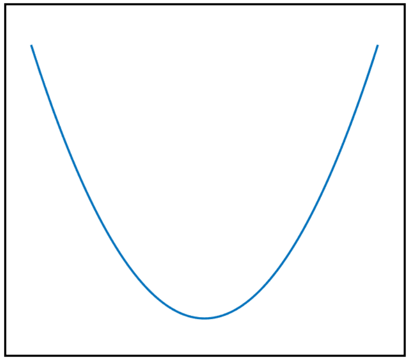
$$\min f(x)$$

$$x_{k+1} = x_k - h\nabla f(x_k)$$

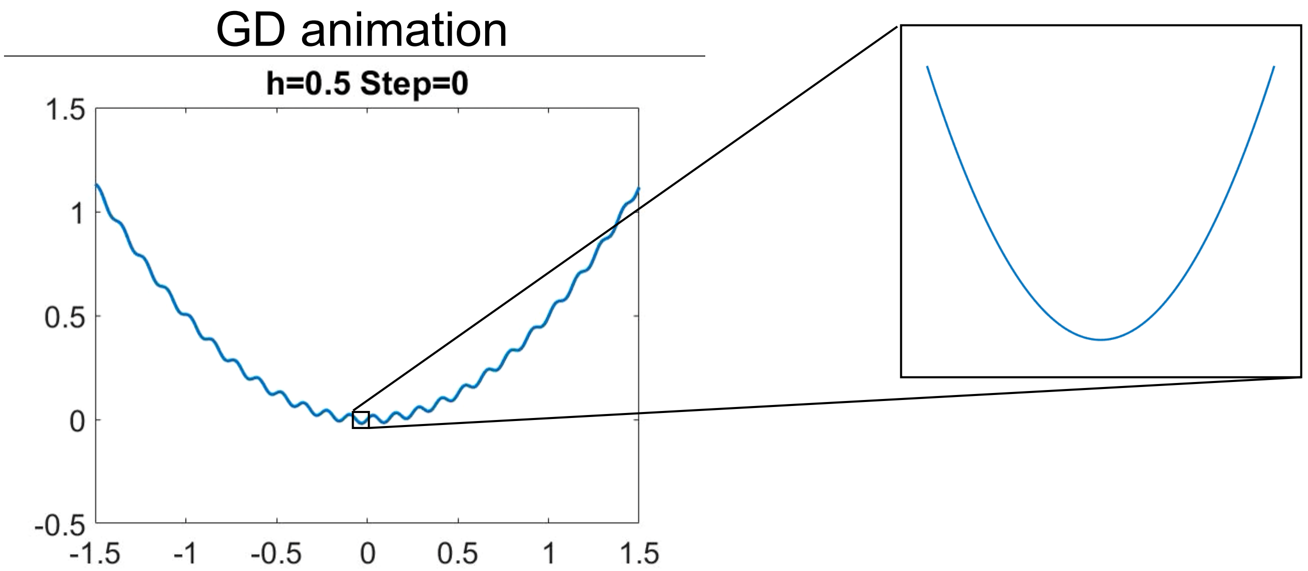
Gradient Descent



1-2. Introduction: GD with `large' LR for multiscale objective -- phenomenology

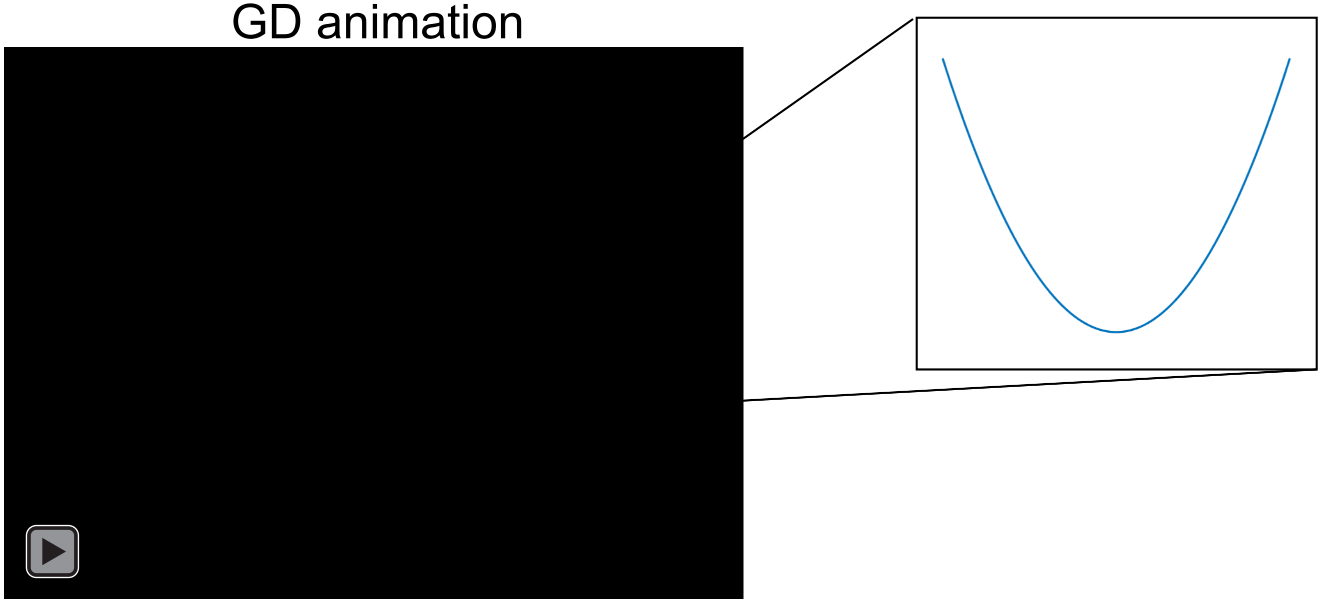


1-2. Introduction: GD with `large' LR for multiscale objective -- phenomenology

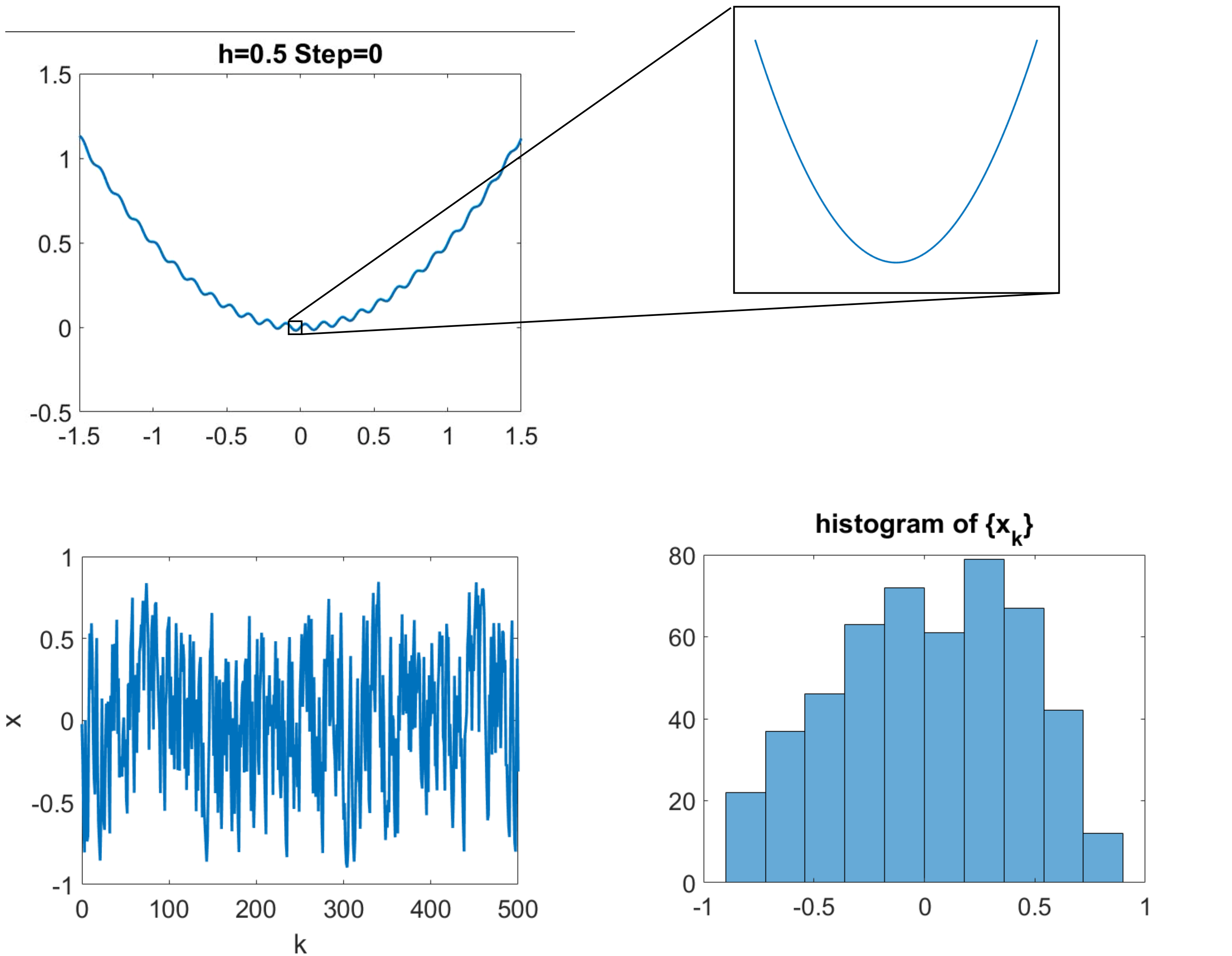


1-2. Introduction: GD with `large' LR for multiscale objective -- phenomenology

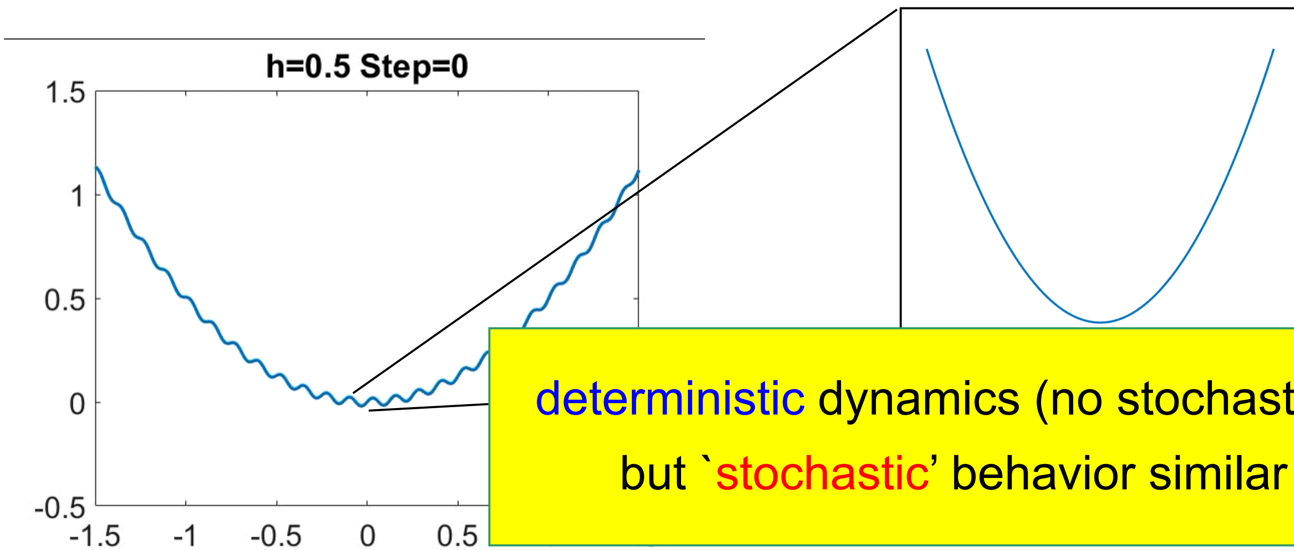
GD animation



1-2. Introduction: GD with `large' LR for multiscale objective -- phenomenology

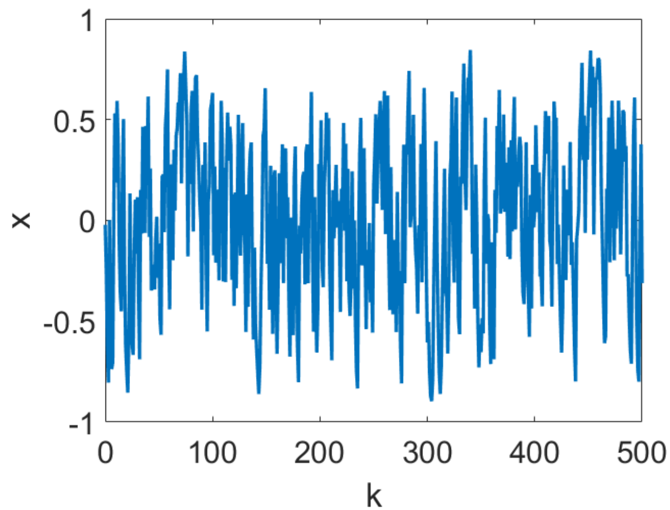


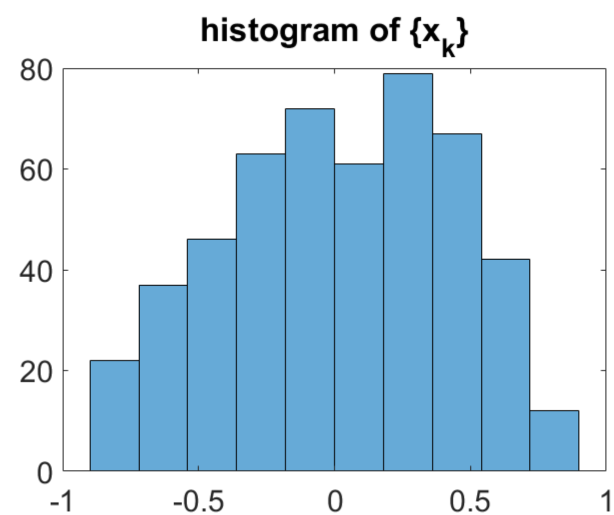
1-2. Introduction: GD with `large' LR for multiscale objective -- phenomenology



deterministic dynamics (no stochastic gradient)

but `stochastic' behavior similar to SGD





2-1. Theory: the setup: multiscale objective function

multiscale objective +
deterministic GD with large LR

chaos $\Rightarrow$

\`**stochastic**’
behaviors

multiscale objective +

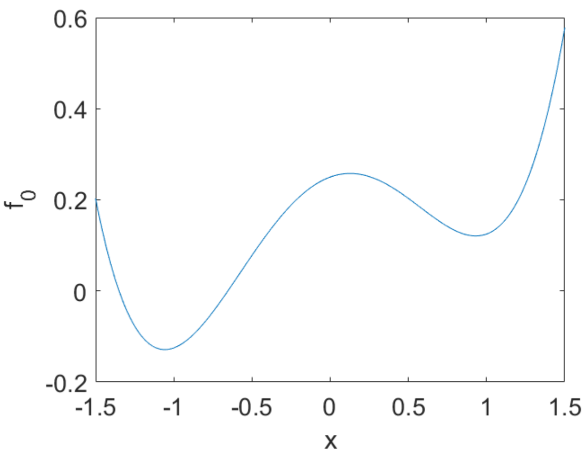
deterministic GD with large LR

multiscale objective fct. $f : \mathbb{R}^d \to \mathbb{R}$. $\quad f(x) := f_0(x) + f_{1,\epsilon}(x)$

multiscale objective +
deterministic GD with large LR

macro

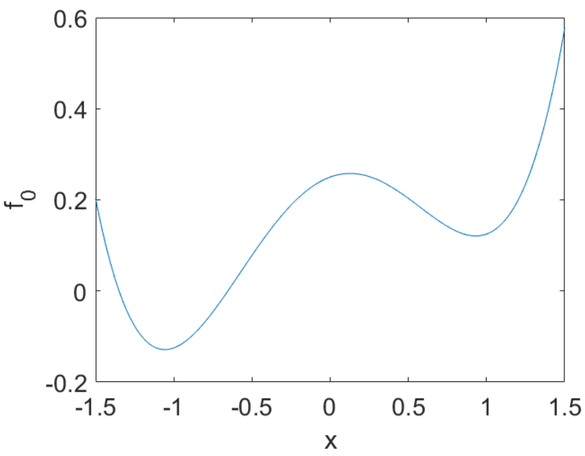
multiscale objective fct. $f : \mathbb{R}^d \to \mathbb{R}$. $\quad f(x) := \boxed{f_0(x)} + f_{1,\epsilon}(x)$



multiscale objective +
deterministic GD with large LR

macro micro $\epsilon \ll 1$

$$\text{multiscale objective fct. } f : \mathbb{R}^d \to \mathbb{R}. \quad f(x) := \boxed{f_0(x)} + \boxed{f_{1,\epsilon}(x)}$$



multiscale objective +

deterministic GD with large LR

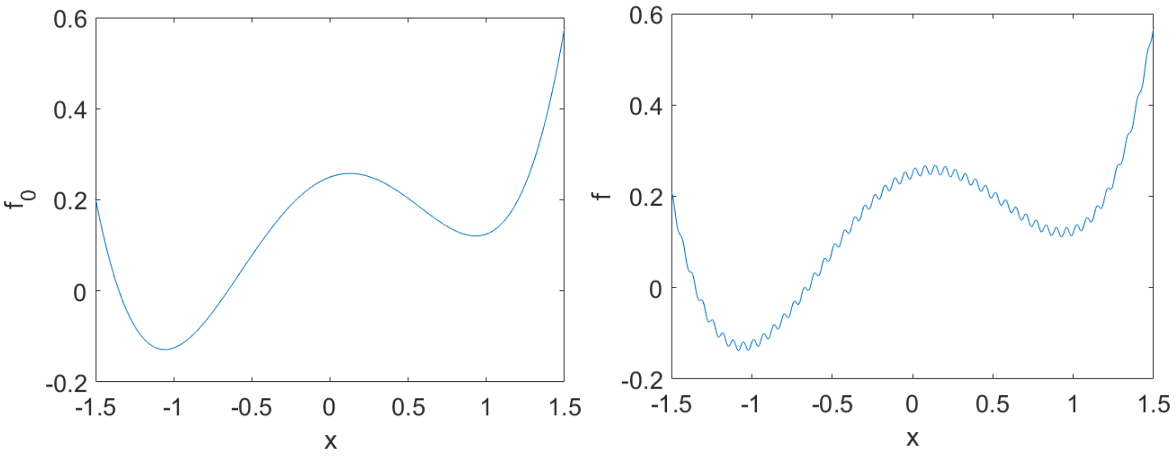
micro

$\epsilon \ll 1$

multiscale objective fct. $f : \mathbb{R}^d \to \mathbb{R}$. $\quad f(x) := f_0(x) + \boxed{f_{1,\epsilon}(x)}$

EX $\quad f_{1,\epsilon} := \epsilon f_1\left(\frac{x}{\epsilon}\right)$

f_1 periodic



multiscale objective +
deterministic GD with large LR

micro

$\epsilon \ll 1$

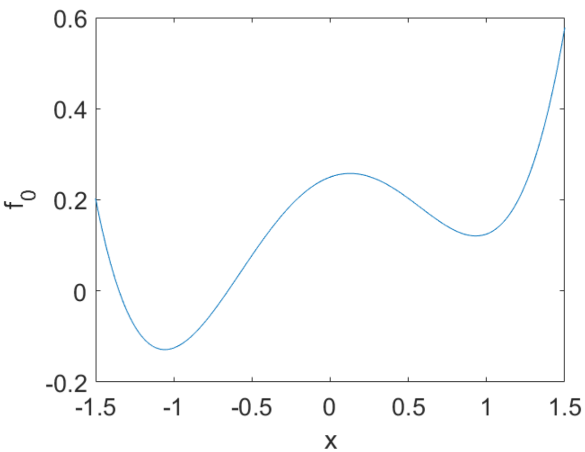
multiscale objective fct. $f : \mathbb{R}^d \to \mathbb{R}$. $\quad f(x) := f_0(x) + \boxed{f_{1,\epsilon}(x)}$

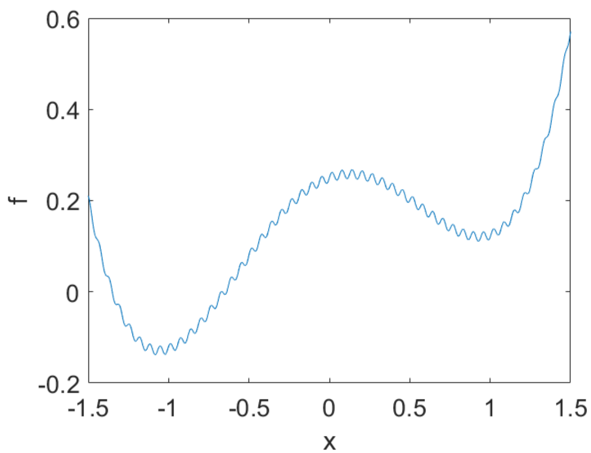
EX $\quad f_{1,\epsilon} := \epsilon f_1\left(\frac{x}{\epsilon}\right)$

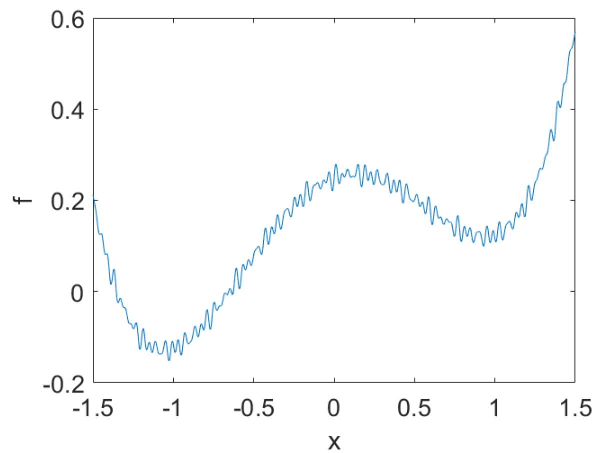
f_1 periodic

EX $\quad f_{1,\epsilon} := \epsilon F_1\left(\frac{\omega_1 x}{\epsilon}, \frac{\omega_2 x}{\epsilon}, \ldots, \frac{\omega_N x}{\epsilon}\right)$

F_1 1-periodic in each argument







multiscale objective +
deterministic GD with large LR

micro

$\epsilon \ll 1$

multiscale objective fct. $f : \mathbb{R}^d \to \mathbb{R}$. $f(x) := f_0(x) + \boxed{f_{1,\epsilon}(x)}$

EX $f_{1,\epsilon} := \epsilon f_1\left(\frac{x}{\epsilon}\right)$

f_1 periodic

EX $f_{1,\epsilon} := \epsilon F_1\left(\frac{\omega_1 x}{\epsilon}, \frac{\omega_2 x}{\epsilon}, \ldots, \frac{\omega_N x}{\epsilon}\right)$

F_1 1-periodic in each argument

more general:

Condition 1. *$\exists$ bounded random variable ζ with $P(\zeta = 0) < 1$, s.t. for any bounded rectangle $\Gamma \subset \mathbb{R}^d$ with area $|\Gamma| > 0$, $-\nabla f_{1,\epsilon}(U_\Gamma) \xrightarrow{w} \zeta$ when $\epsilon \to 0$. Assume without loss of generality that $\mathbb{E}\zeta = 0$ (nonzero mean can be absorbed into f_0).*

Condition 2. *$\epsilon \nabla^2 f_{1,\epsilon}$ is uniformly bounded as $\epsilon \to 0$, and $\exists m \in \mathbb{R}$, s.t. for any bounded rectangle $\Gamma \subset \mathbb{R}^d$ whose area $|\Gamma| > 0$, $\mathbb{E}\left[\ln \|\epsilon \nabla^2 f_{1,\epsilon}(U_\Gamma)\|_2\right] \to m$.*

multiscale objective +
deterministic GD with large LR

micro

$\epsilon \ll 1$

multiscale objective fct. $f : \mathbb{R}^d \to \mathbb{R}$. $\quad f(x) := f_0(x) + \boxed{f_{1,\epsilon}(x)}$

EX $f_{1,\epsilon} := \epsilon f_1\left(\frac{x}{\epsilon}\right)$

f_1 periodic

EX $f_{1,\epsilon} := \epsilon F_1\left(\frac{\omega_1 x}{\epsilon}, \frac{\omega_2 x}{\epsilon}, \ldots, \frac{\omega_N x}{\epsilon}\right)$

F_1 1-periodic in each argument

more general:

O(1) 1st-derivative with a weak limit

Condition 1. *$\exists$ bounded random variable ζ with $P(\zeta = 0) < 1$, s.t. for any bounded rectangle $\Gamma \subset \mathbb{R}^d$ with area $|\Gamma| > 0$, $\boxed{-\nabla f_{1,\epsilon}(U_\Gamma) \xrightarrow{w} \zeta}$ when $\epsilon \to 0$. Assume without loss of generality that $\mathbb{E}\zeta = 0$ (nonzero mean can be absorbed into f_0).*

Condition 2. *$\epsilon \nabla^2 f_{1,\epsilon}$ is uniformly bounded as $\epsilon \to 0$, and $\exists m \in \mathbb{R}$, s.t. for any bounded rectangle $\Gamma \subset \mathbb{R}^d$ whose area $|\Gamma| > 0$, $\mathbb{E}\left[\ln \|\epsilon \nabla^2 f_{1,\epsilon}(U_\Gamma)\|_2\right] \to m$.*

multiscale objective +
deterministic GD with large LR

micro

$\epsilon \ll 1$

multiscale objective fct. $f : \mathbb{R}^d \to \mathbb{R}$. $f(x) := f_0(x) + \boxed{f_{1,\epsilon}(x)}$

EX $f_{1,\epsilon} := \epsilon f_1\left(\frac{x}{\epsilon}\right)$

f_1 periodic

EX $f_{1,\epsilon} := \epsilon F_1\left(\frac{\omega_1 x}{\epsilon}, \frac{\omega_2 x}{\epsilon}, \ldots, \frac{\omega_N x}{\epsilon}\right)$

F_1 1-periodic in each argument

more general:

O(1) 1st-derivative with a weak limit

Condition 1. *$\exists$ bounded random variable ζ with $P(\zeta = 0) < 1$, s.t. for any bounded rectangle $\Gamma \subset \mathbb{R}^d$ with area $|\Gamma| > 0$, $-\nabla f_{1,\epsilon}(U_\Gamma) \xrightarrow{w} \zeta$ when $\epsilon \to 0$. Assume without loss of generality that $\mathbb{E}\zeta = 0$ (nonzero mean can be absorbed into f_0).*

Condition 2. *$\epsilon\nabla^2 f_{1,\epsilon}$ is uniformly bounded as $\epsilon \to 0$, and $\exists m \in \mathbb{R}$, s.t. for any bounded rectangle $\Gamma \subset \mathbb{R}^d$ whose area $|\Gamma| > 0$, $\mathbb{E}\left[\ln \|\epsilon\nabla^2 f_{1,\epsilon}(U_\Gamma)\|_2\right] \to m$.*

O(1/ε) 2nd-derivative with a limiting expection

multiscale objective +
deterministic GD with large LR

$$f(x) := f_0(x) + f_{1,\epsilon}(x)$$
$$\text{e.g., } f_{1,\epsilon} = \epsilon f_1(x/\epsilon)$$

$$x \mapsto x - \eta \nabla f(x), \quad \text{i.e.} \quad x \mapsto x - \eta(\nabla f_0(x) + \nabla f_{1,\epsilon}(x))$$

multiscale objective +
deterministic GD with large LR

$$f(x) := f_0(x) + f_{1,\epsilon}(x)$$
$$\text{e.g., } f_{1,\epsilon} = \epsilon f_1(x/\epsilon)$$

$$x \mapsto x - \eta \nabla f(x), \quad \text{i.e.} \quad x \mapsto x - \eta(\nabla f_0(x) + \nabla f_{1,\epsilon}(x))$$

small LR?

multiscale objective +
deterministic GD with large LR

$$f(x) := f_0(x) + f_{1,\epsilon}(x)$$
$$\text{e.g., } f_{1,\epsilon} = \epsilon f_1(x/\epsilon)$$

$$x \mapsto x - \eta \nabla f(x), \quad \text{i.e.} \quad x \mapsto x - \eta(\nabla f_0(x) + \nabla f_{1,\epsilon}(x))$$

small LR? $\eta \ll 1/L$ L: Lipschitz const. of ∇f, $L = \mathcal{O}(1/\epsilon)$

i.e., $\eta \ll \epsilon$ resolve the small scale → conv. to local min

multiscale objective +
deterministic GD with large LR

$$f(x) := f_0(x) + f_{1,\epsilon}(x)$$
$$\text{e.g., } f_{1,\epsilon} = \epsilon f_1(x/\epsilon)$$

$$x \mapsto x - \eta \nabla f(x), \quad \text{i.e.} \quad x \mapsto x - \eta(\nabla f_0(x) + \nabla f_{1,\epsilon}(x))$$

small LR? $\eta \ll 1/L$ L: Lipschitz const. of ∇f, $L = \mathcal{O}(1/\epsilon)$

i.e., $\eta \ll \epsilon$ resolve the small scale → conv. to local min

bad LR? $\eta \gg 1$ can't even resolve the large scale

i.e., $x \mapsto x - \eta \nabla f_0(x)$ is unstable

2-1. Theory: the setup: large Learning Rate

multiscale objective +
deterministic GD with large LR

$$f(x) := f_0(x) + f_{1,\epsilon}(x)$$
$$\text{e.g., } f_{1,\epsilon} = \epsilon f_1(x/\epsilon)$$

$$x \mapsto x - \eta \nabla f(x), \quad \text{i.e.} \quad x \mapsto x - \eta(\nabla f_0(x) + \nabla f_{1,\epsilon}(x))$$

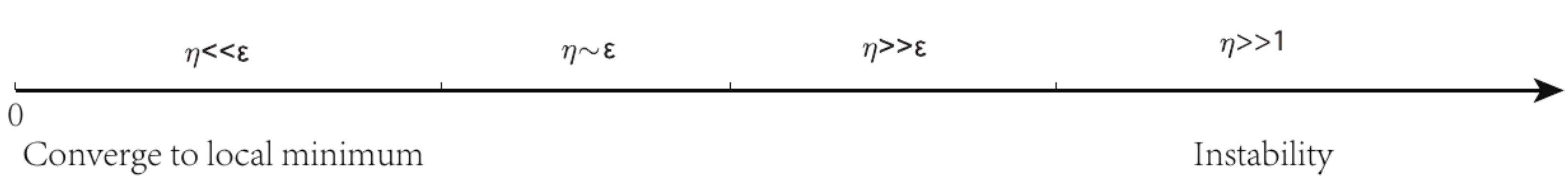
small LR? $\eta \ll 1/L$ L: Lipschitz const. of ∇f, $L = \mathcal{O}(1/\epsilon)$

i.e., $\eta \ll \epsilon$ resolve the small scale → conv. to local min

bad LR? $\eta \gg 1$ can't even resolve the large scale

i.e., $x \mapsto x - \eta \nabla f_0(x)$ is unstable

in-between…



2-1. Theory: the setup: large Learning Rate

multiscale objective + deterministic GD with large LR

$$f(x) := f_0(x) + f_{1,\epsilon}(x)$$
$$\text{e.g.,}\ f_{1,\epsilon} = \epsilon f_1(x/\epsilon)$$

$$x \mapsto x - \eta \nabla f(x), \quad \text{i.e.} \quad x \mapsto x - \eta(\nabla f_0(x) + \nabla f_{1,\epsilon}(x))$$

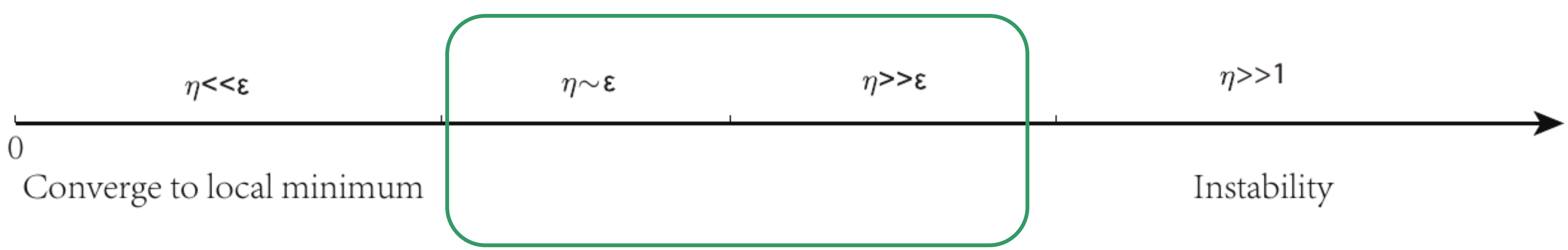
small LR? $\eta \ll 1/L$ $\quad L$: Lipschitz const. of ∇f, $L = \mathcal{O}(1/\epsilon)$

i.e., $\eta \ll \epsilon$ $\quad$ resolve the small scale → conv. to local min

bad LR? $\eta \gg 1$ $\quad$ can't even resolve the large scale

i.e., $x \mapsto x - \eta \nabla f_0(x)$ is unstable

in-between…



2-1. Theory: the setup: large Learning Rate

multiscale objective + deterministic GD with large LR

$$f(x) := f_0(x) + f_{1,\epsilon}(x)$$

$$\text{e.g., } f_{1,\epsilon} = \epsilon f_1(x/\epsilon)$$

$$x \mapsto x - \eta \nabla f(x), \quad \text{i.e.} \quad x \mapsto x - \eta(\nabla f_0(x) + \nabla f_{1,\epsilon}(x))$$

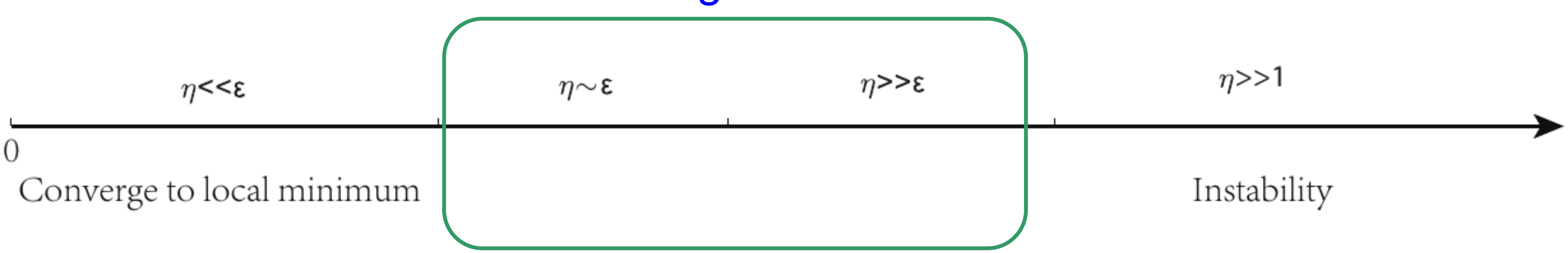
small LR? $\eta \ll 1/L$ L: Lipschitz const. of ∇f, $L = \mathcal{O}(1/\epsilon)$

i.e., $\eta \ll \epsilon$ resolve the small scale → conv. to local min

bad LR? $\eta \gg 1$ can't even resolve the large scale

i.e., $x \mapsto x - \eta \nabla f_0(x)$ is unstable

in-between… large LR



multiscale objective + deterministic GD with large LR

$$f(x) := f_0(x) + f_{1,\epsilon}(x)$$
$$\text{e.g., } f_{1,\epsilon} = \epsilon f_1(x/\epsilon)$$

$$x \mapsto x - \eta \nabla f(x), \quad \text{i.e.} \quad x \mapsto x - \eta(\nabla f_0(x) + \nabla f_{1,\epsilon}(x))$$

small LR? $\eta \ll 1/L$ L: Lipschitz const. of ∇f, $L = \mathcal{O}(1/\epsilon)$

i.e., $\eta \ll \epsilon$ resolve the small scale → conv. to local min

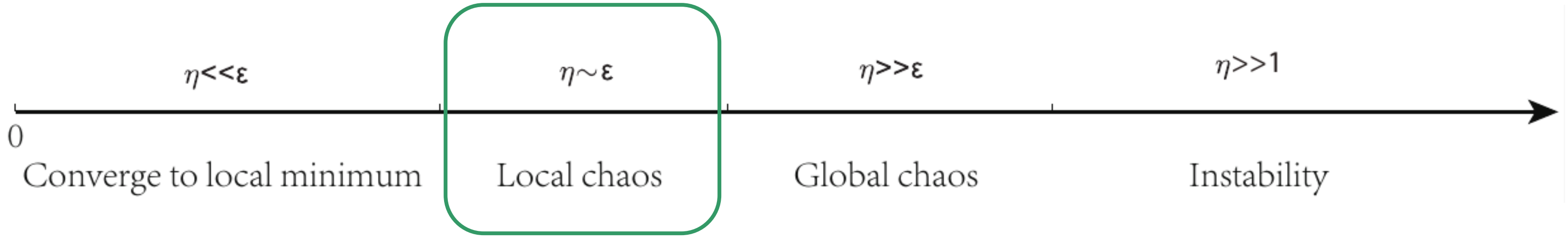
bad LR? $\eta \gg 1$ can't even resolve the large scale

i.e., $x \mapsto x - \eta \nabla f_0(x)$ is unstable

in-between… large LR

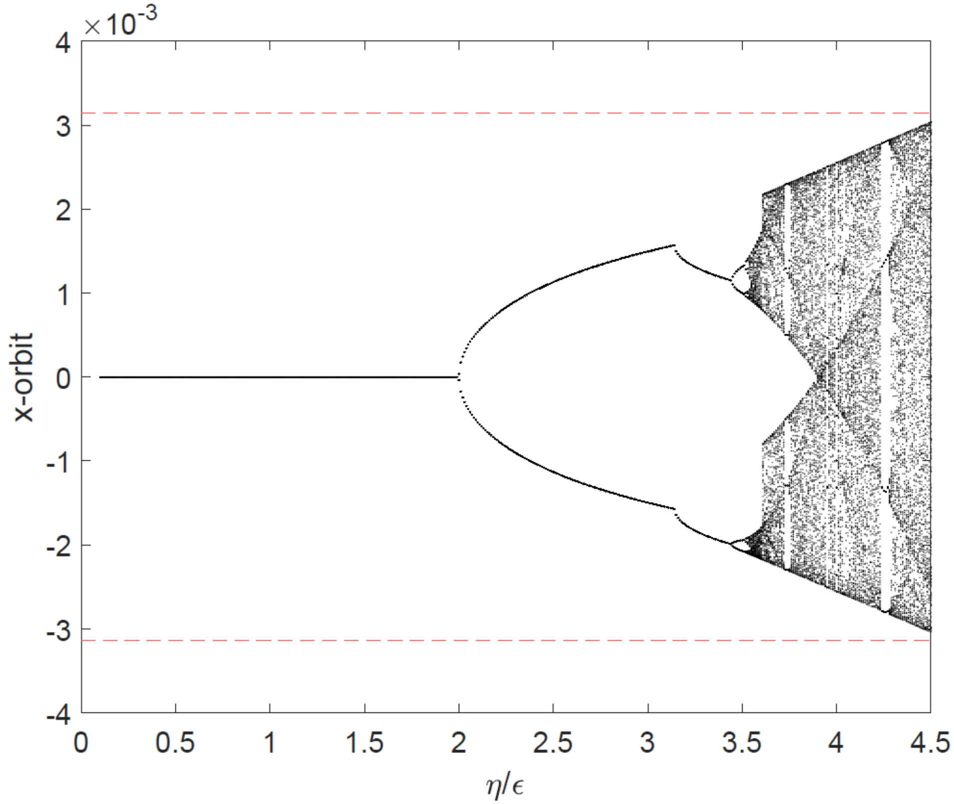
2-2. Theory: the local chaos regime

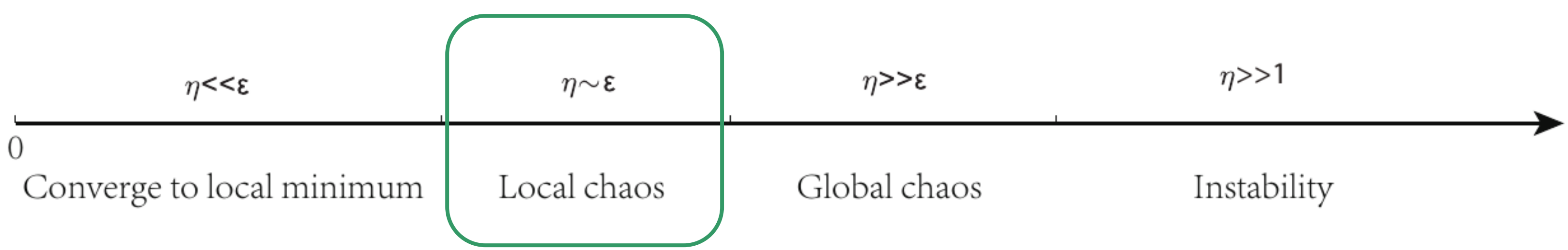
deterministic GD map $\varphi(x) := x - \eta(\nabla f_0(x) + \nabla f_{1,\epsilon}(x))$



deterministic GD map $\varphi(x) := x - \eta(\nabla f_0(x) + \nabla f_{1,\epsilon}(x))$

transition into chaos via period doublings

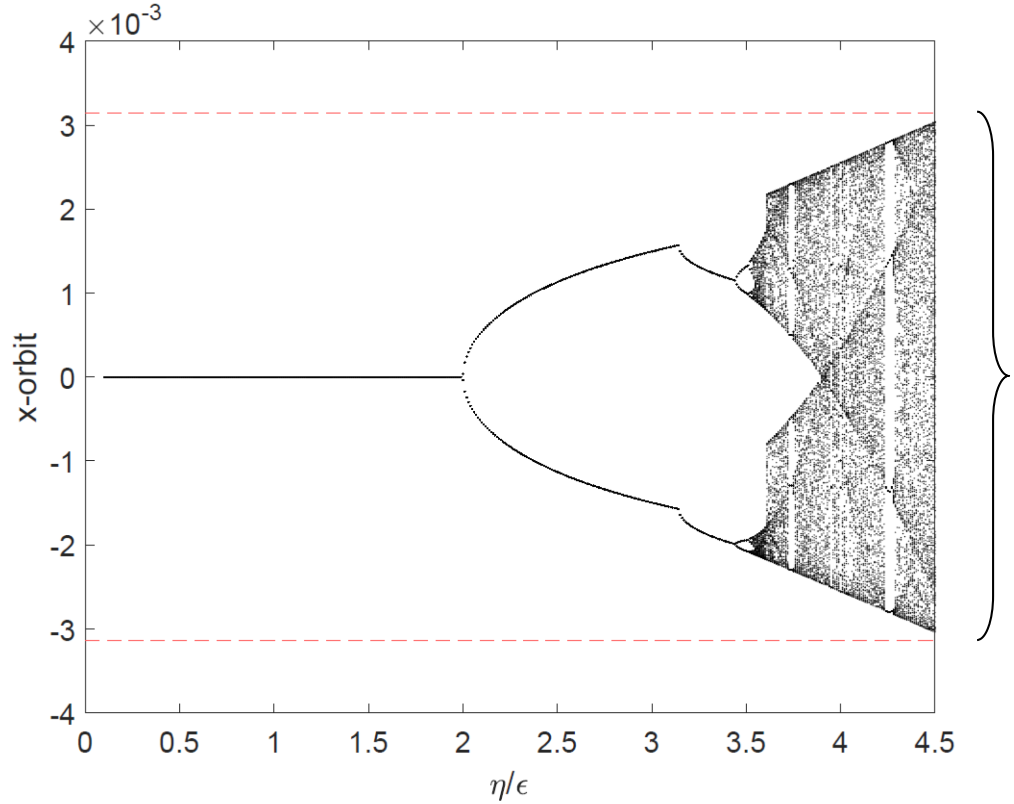




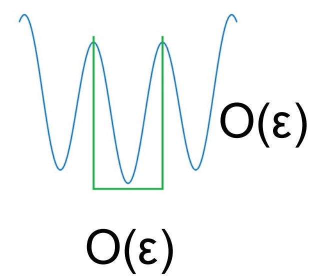
2-2. Theory: the local chaos regime

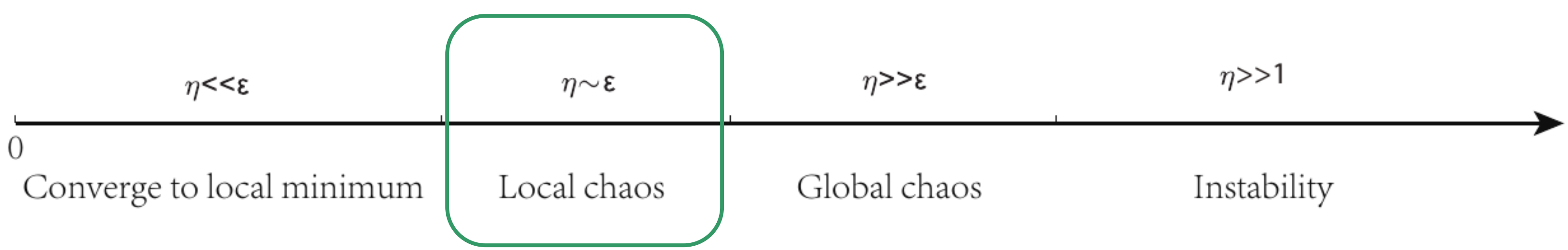
deterministic GD map $\varphi(x) := x - \eta(\nabla f_0(x) + \nabla f_{1,\epsilon}(x))$

transition into chaos via period doublings



x stays within a microscopic potential well



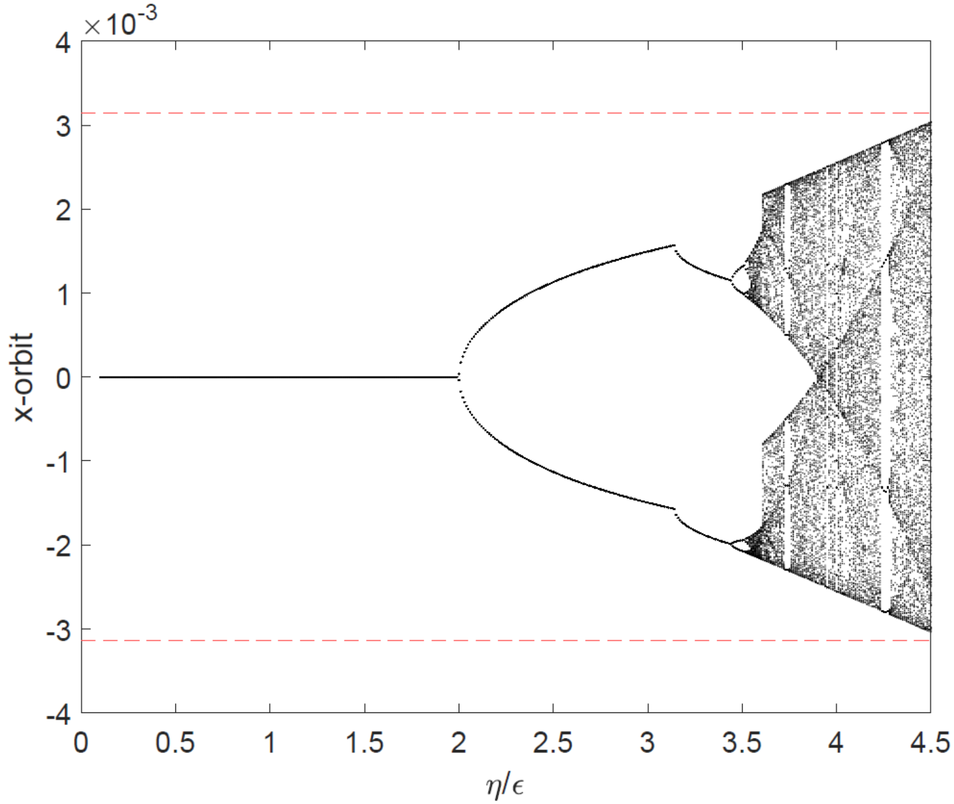


2-2. Theory: the local chaos regime

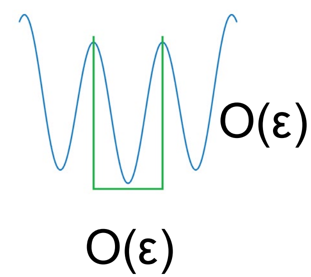
deterministic GD map $\varphi(x) := x - \eta(\nabla f_0(x) + \nabla f_{1,\epsilon}(x))$

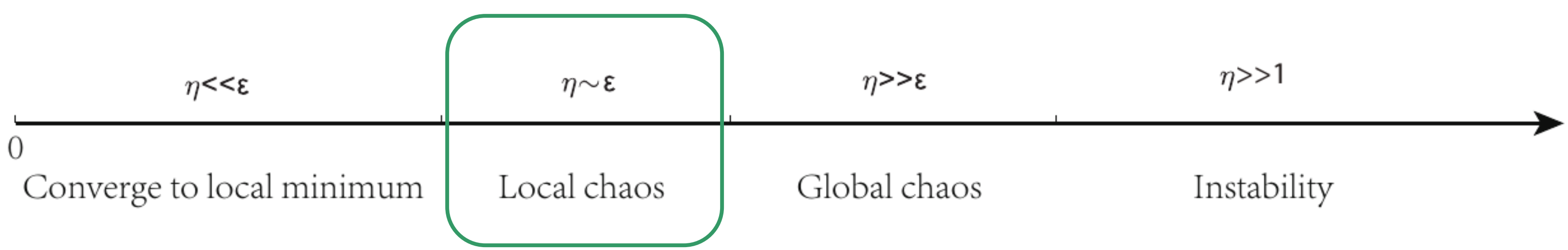
unimodal map within a well

transition into chaos via period doublings



x stays within a microscopic potential well



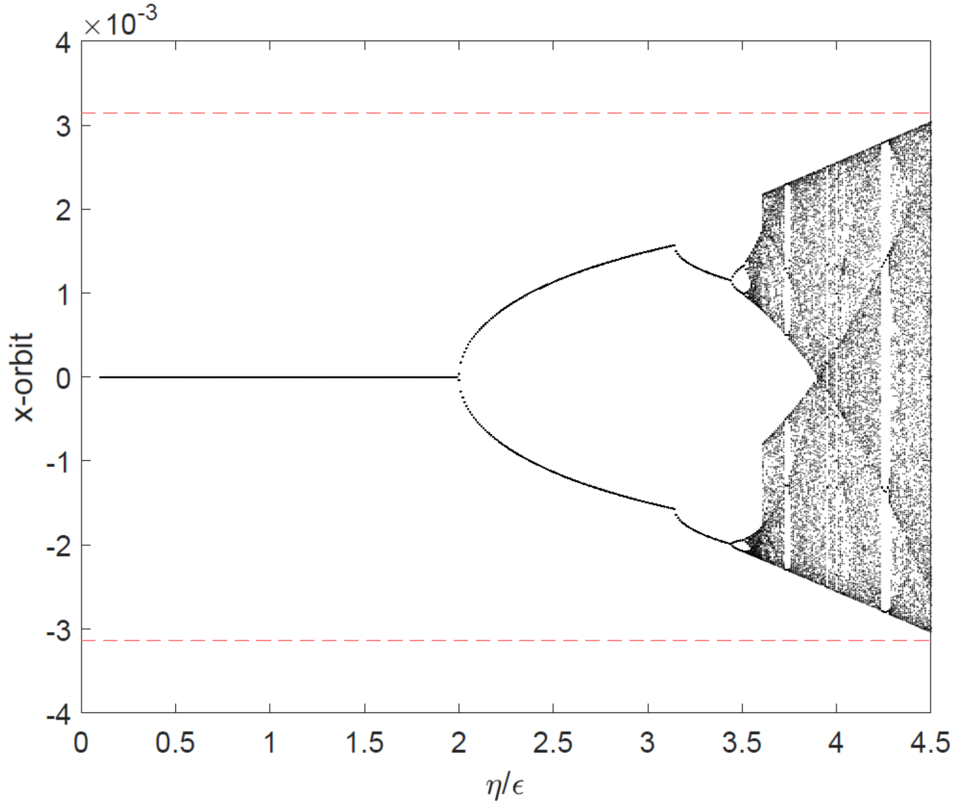


2-2. Theory: the local chaos regime

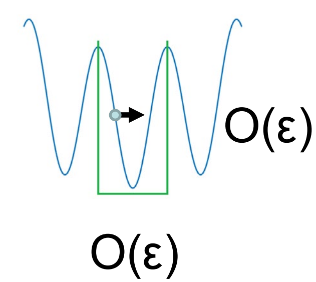
deterministic GD map $\boxed{\varphi(x) := x - \eta(\nabla f_0(x) + \nabla f_{1,\epsilon}(x))}$

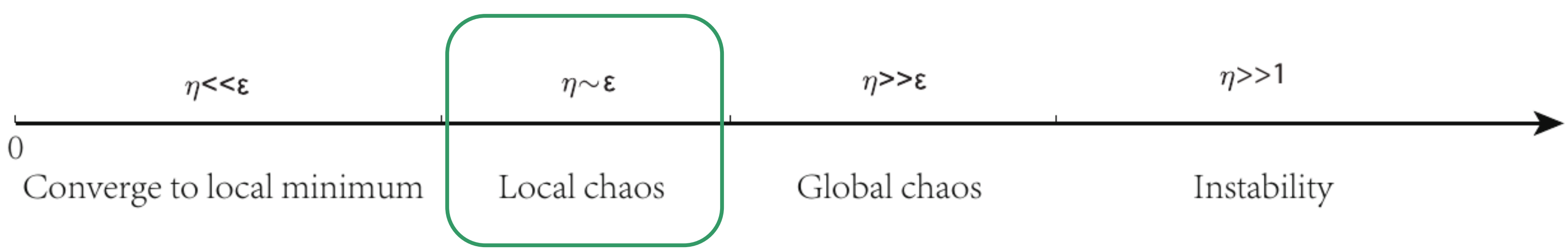
unimodal map within a well

transition into chaos via period doublings



x stays within a microscopic potential well



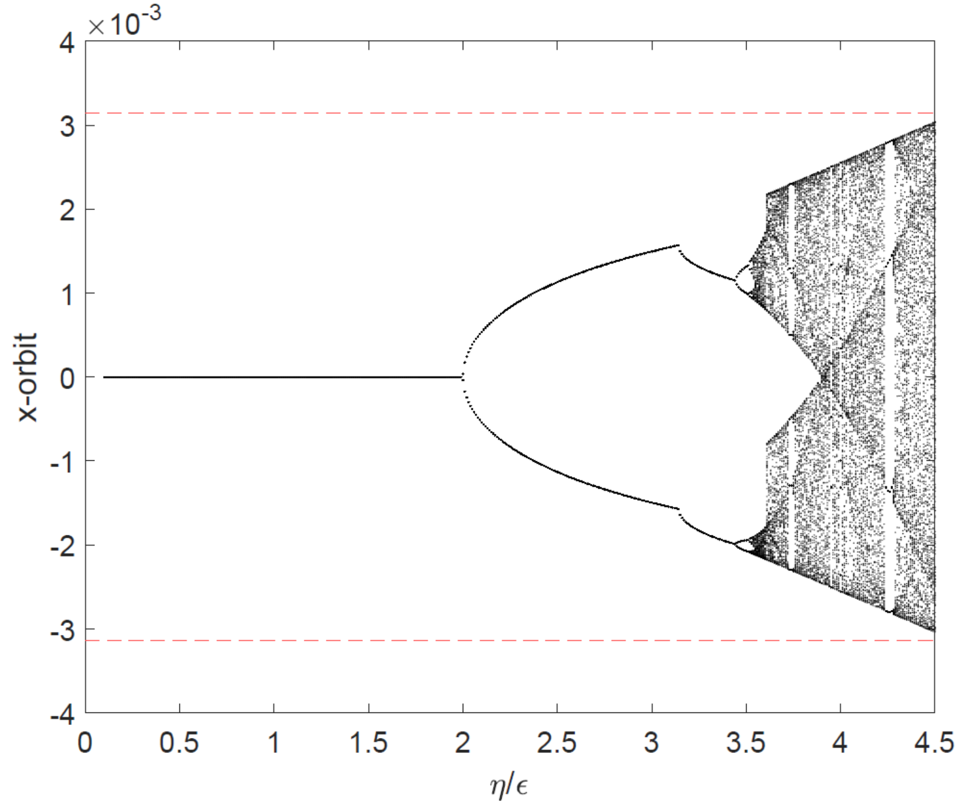


2-2. Theory: the local chaos regime

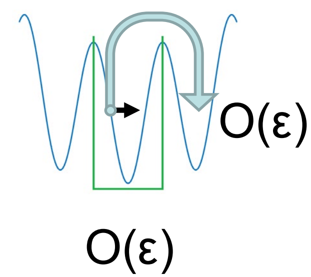
deterministic GD map $\varphi(x) := x - \eta(\nabla f_0(x) + \nabla f_{1,\epsilon}(x))$

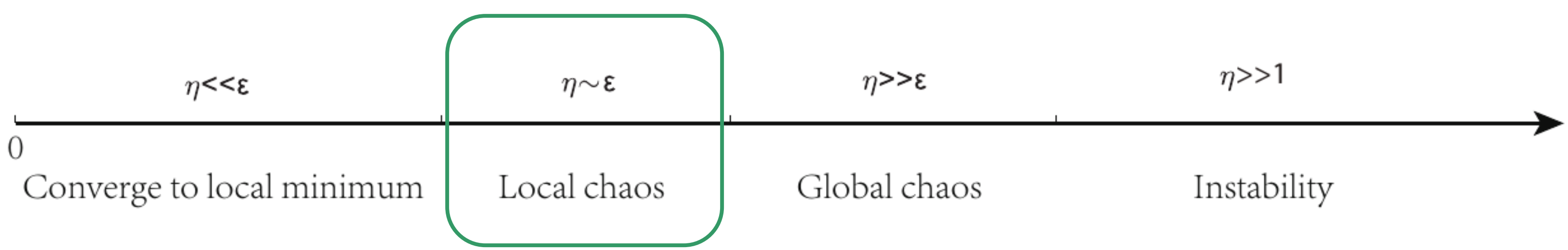
unimodal map within a well

transition into chaos via period doublings



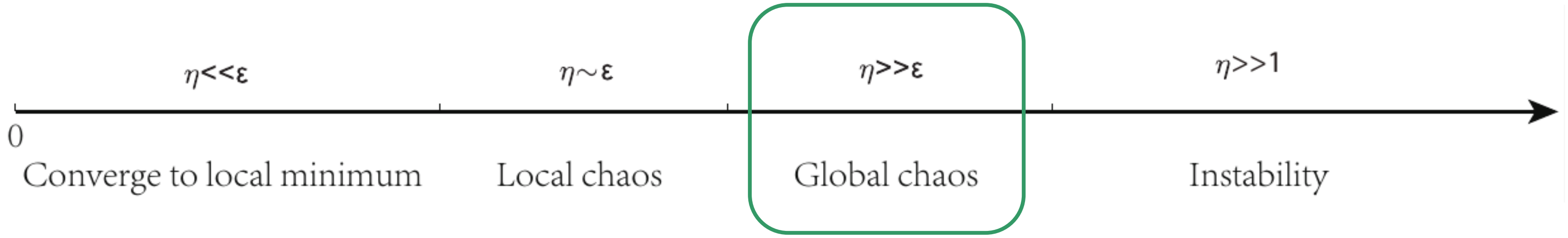
x stays within a microscopic potential well





deterministic GD map $\varphi(x) := x - \eta(\nabla f_0(x) + \nabla f_{1,\epsilon}(x))$

η further increases (i.e. the ε→0 regime)



2-3. Theory: the global chaos regime

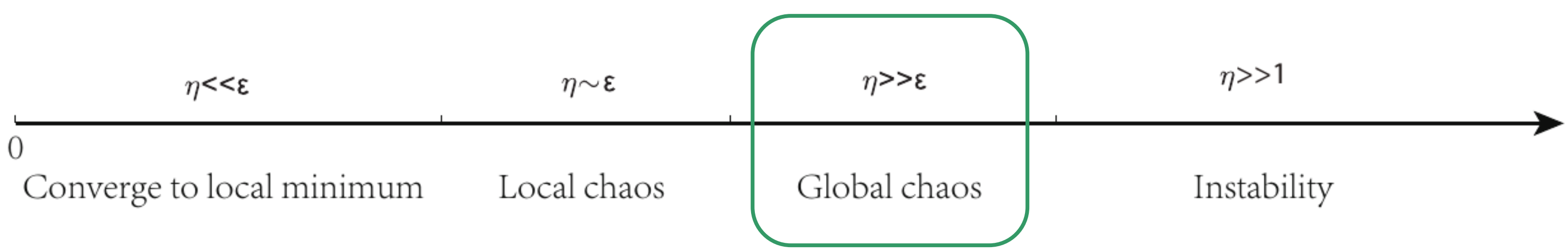
deterministic GD map $\varphi(x) := x - \eta(\nabla f_0(x) + \nabla f_{1,\epsilon}(x))$

η further increases (i.e. the ε→0 regime)

- f_0 strongly convex, C^2, L-smooth →

limiting distribution of φ iterates $\overset{\text{small } \varepsilon}{\approx}$ nearly Gibbs

$$\boxed{\frac{1}{Z}\exp\left(-\frac{2f_0(x)}{\eta\sigma^2}\right)dx + \mathcal{O}(\eta^2)}$$



2-3. Theory: the global chaos regime

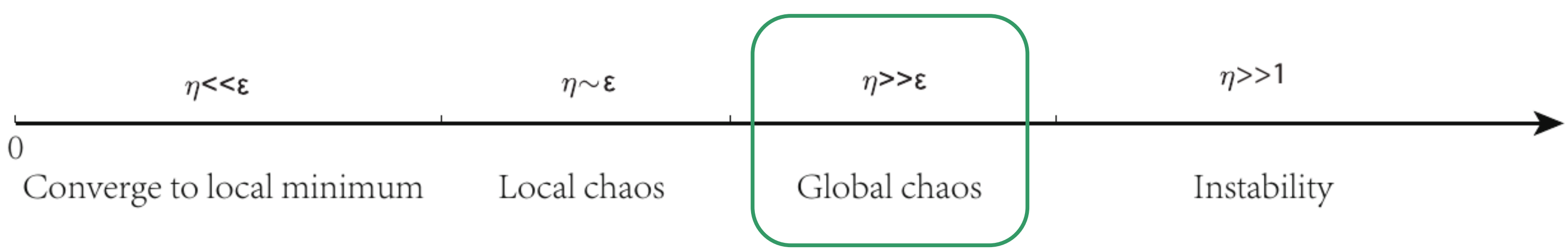
deterministic GD map $\varphi(x) := x - \eta(\nabla f_0(x) + \nabla f_{1,\epsilon}(x))$

η further increases (i.e. the ε→0 regime)

limiting distribution: nearly Gibbs

(f_0 strongly convex)

$$\frac{1}{Z}\exp\left(-\frac{2f_0(x)}{\eta\sigma^2}\right)dx + \mathcal{O}(\eta^2)$$



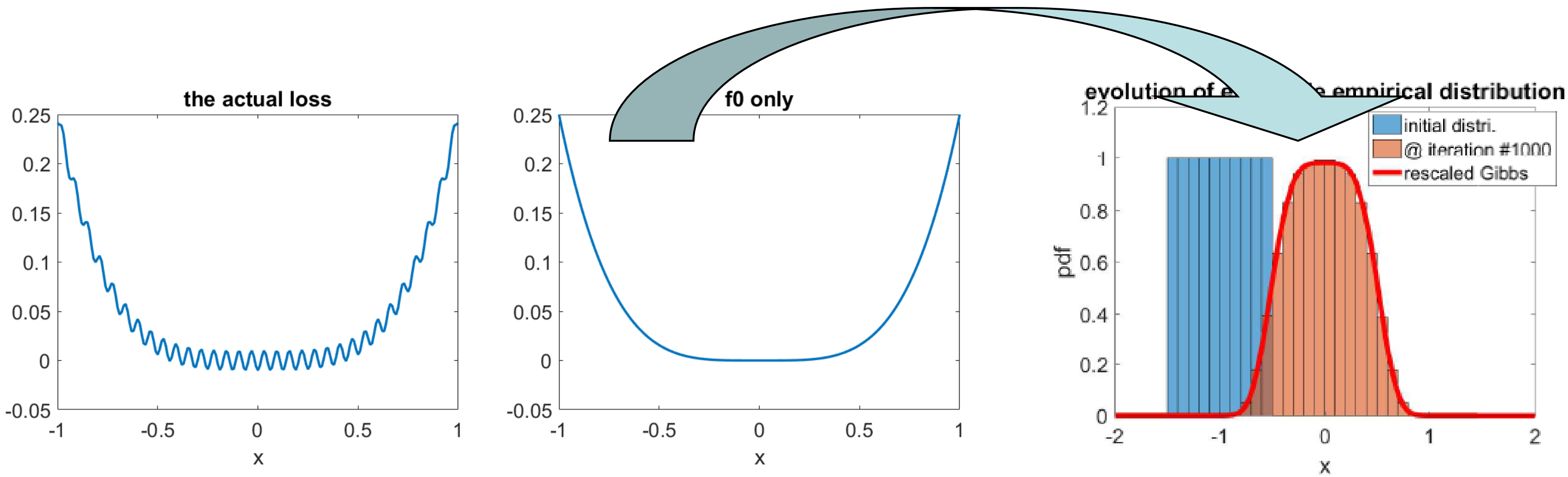
deterministic GD map $\varphi(x) := x - \eta(\nabla f_0(x) + \nabla f_{1,\epsilon}(x))$

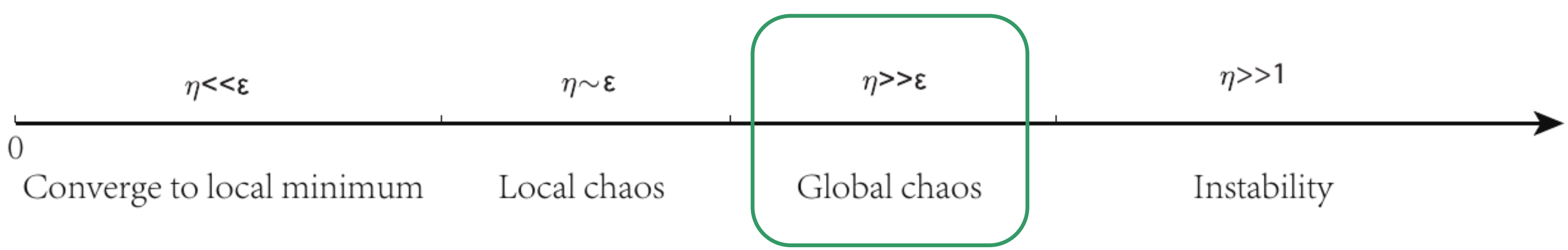
η further increases (i.e. the ε→0 regime)

limiting distribution: nearly Gibbs

(f_0 strongly convex)

$$\frac{1}{Z}\exp\left(-\frac{2f_0(x)}{\eta\sigma^2}\right)dx + \mathcal{O}(\eta^2)$$





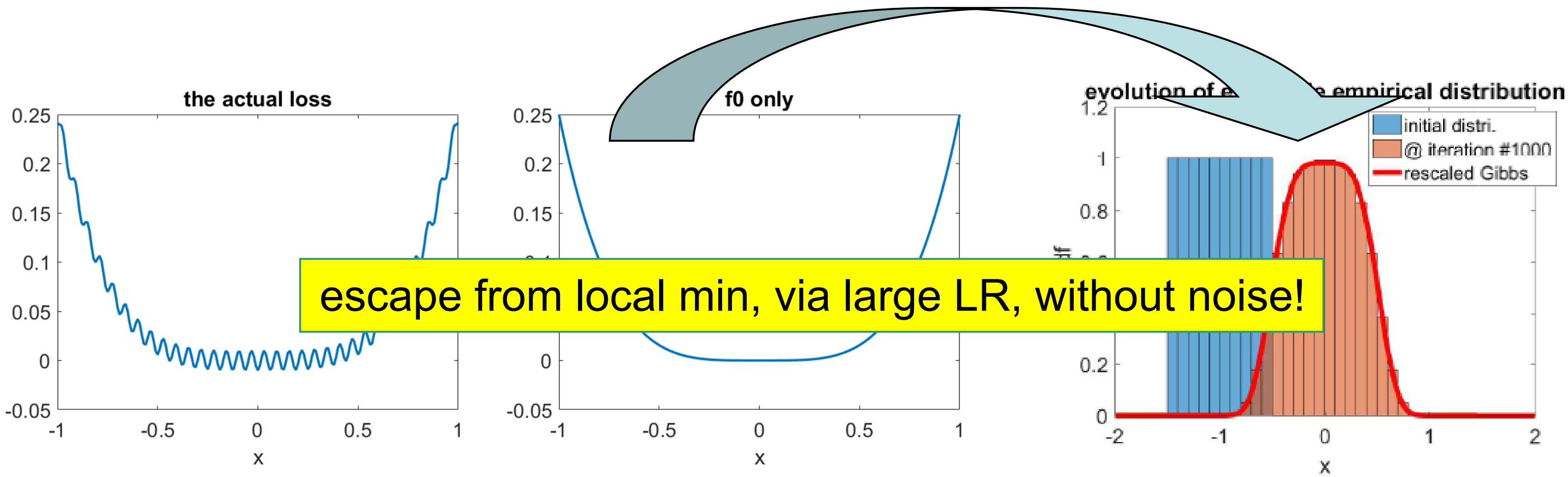
2-3. Theory: the global chaos regime

deterministic GD map $\varphi(x) := x - \eta(\nabla f_0(x) + \nabla f_{1,\epsilon}(x))$

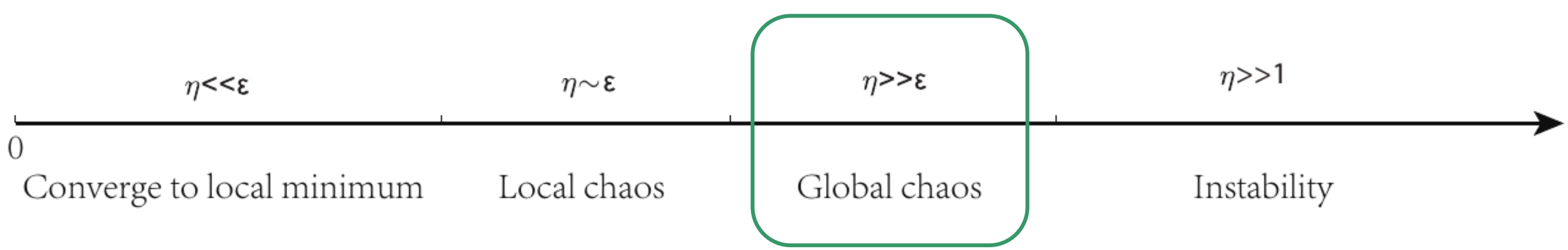
η further increases (i.e. the ε→0 regime) limiting distribution: nearly Gibbs

(f$_0$ strongly convex)

$$\frac{1}{Z}\exp\left(-\frac{2f_0(x)}{\eta\sigma^2}\right)dx + \mathcal{O}(\eta^2)$$



escape from local min, via large LR, without noise!



2-5. Theory: the global chaos regime: non-convex macroscale

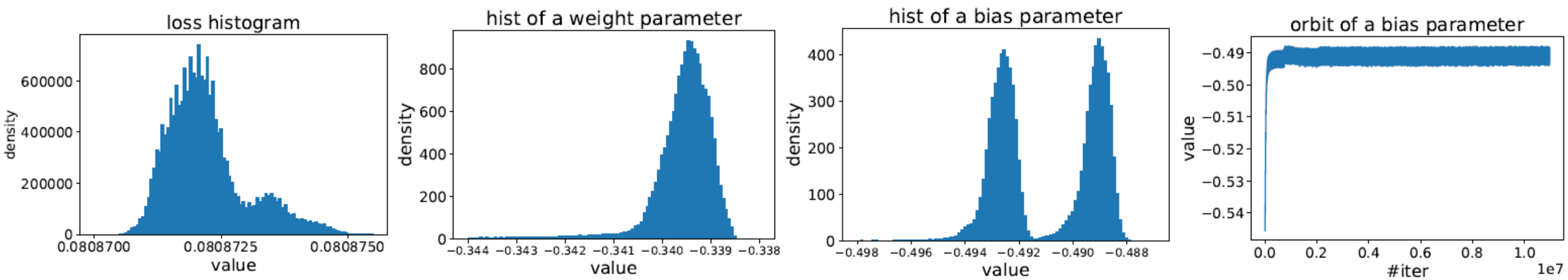
deterministic GD map $$\varphi(x) := x - \eta(\nabla f_0(x) + \nabla f_{1,\epsilon}(x))$$

previously discussed (f_0 strongly convex): $$\frac{1}{Z}\exp\left(-\frac{2f_0(x)}{\eta\sigma^2}\right)dx + \mathcal{O}(\eta^2)$$

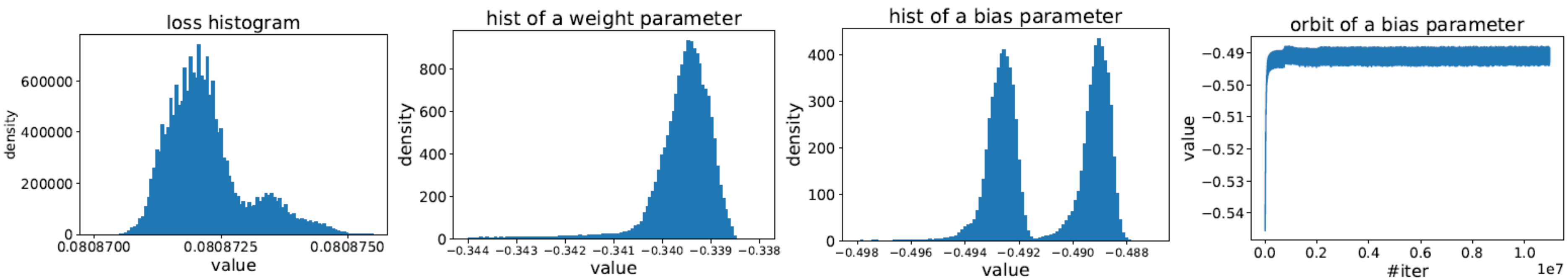
? non-convex f_0

❓ Is f for real problems multiscale?

EX 5-16-2 FF neural network, regression, UCI Airfoil Self-Noise dataset, large LR

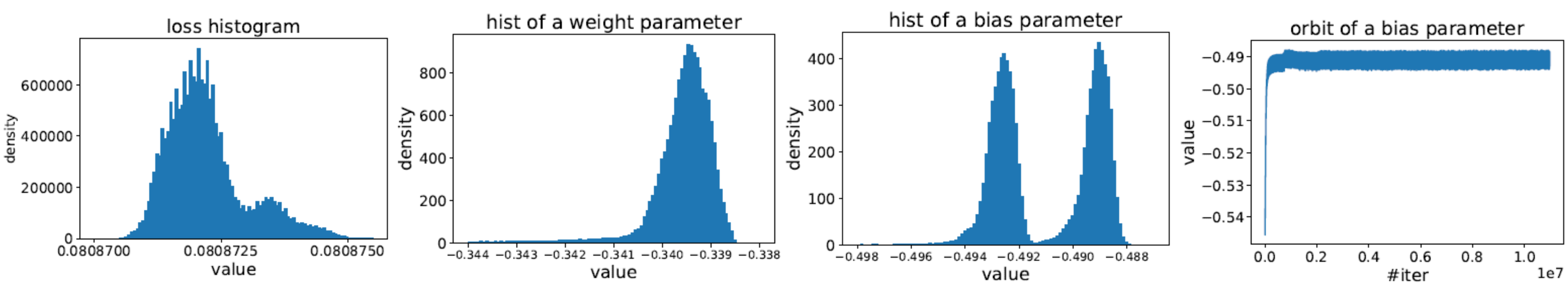


EX 5-16-2 FF neural network, regression, UCI Airfoil Self-Noise dataset, large LR



Theory multiscale data (mean + fluctuation) → multiscale loss?

EX 5-16-2 FF neural network, regression, UCI Airfoil Self-Noise dataset, large LR



Theory multiscale data (mean + fluctuation) → multiscale loss?

2 layer, periodic activation

chaotic dynamics can help deterministic GD
escape **microscopic** local minima
as it ~~optimizes the loss~~ samples a distribution

Summary

chaotic dynamics can help deterministic GD
escape **microscopic** local minima
as it ~~optimizes the loss~~ samples a distribution

? why this matters

Summary

chaotic dynamics can help deterministic GD
escape **microscopic** local minima
as it ~~optimizes the loss~~ samples a distribution

? why this matters

deeper minimum → better training (& thus test) accuracies

Thank you for your attention and feedback!

Support:

NSF DMS-1847802, ECCS-1936776

GT Cullen-Peck Scholarship

Emory-GT AI.Humanity Award

itsdynamical

MoleiTaoMath

SCAN ME

SCAN ME