Properties of Fixed Points of Generalised Extra Gradient Methods Applied to Min-Max Problems

Philipp Braun

Australian National University, CIICADA Lab

Acknowledgments

Amir Ali Farzin

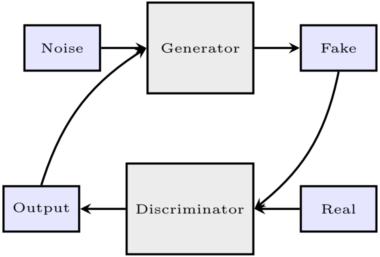
Yuen-Man Pun

Iman Shames

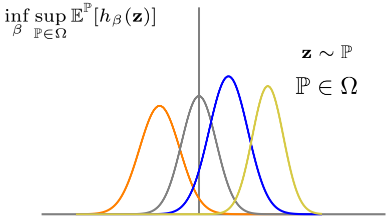
Amir Ali Farzin, Yuen-Man Pun, Philipp Braun, and Iman Shames. *"Properties of Fixed Points of Generalised Extra Gradient Methods Applied to Min-Max Problems."* IEEE Control Systems Letters (2025).

$$\min_x \max_y f(x, y),$$

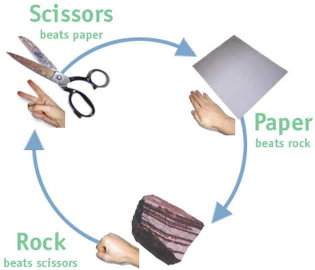
- Generative Adversarial Networks



- Distributionally Robust Optimization



- Non-cooperative Games



Fixed Point Iterations & Dynamical Systems

Gradient Descent Ascent (GDA):

$$x_{k+1} = x_k - \frac{\eta}{\tau}\nabla_x f(x_k, y_k)$$
$$y_{k+1} = y_k + \eta\nabla_y f(x_k, y_k)$$

(step sizes τ, η)

Fixed Point Iterations & Dynamical Systems

Gradient Descent Ascent (GDA):

$$x_{k+1} = x_k - \frac{\eta}{\tau} \nabla_x f(x_k, y_k)$$
$$y_{k+1} = y_k + \eta \nabla_y f(x_k, y_k)$$

(step sizes τ, η)

Discrete-time dynamical system:

$$z_{k+1} = w(z_k)$$

Fixed Point Iterations & Dynamical Systems

Gradient Descent Ascent (GDA):

$$x_{k+1} = x_k - \frac{\eta}{\tau}\nabla_x f(x_k, y_k)$$
$$y_{k+1} = y_k + \eta\nabla_y f(x_k, y_k)$$

(step sizes τ, η)

Discrete-time dynamical system:

$$z_{k+1} = w(z_k)$$

If we define

$$z = \begin{bmatrix} x \\ y \end{bmatrix}, \ \Lambda_\tau = \begin{bmatrix} \frac{1}{\tau}I & 0 \\ 0 & I \end{bmatrix}$$
$$F(z) = \begin{bmatrix} \nabla_x f(z) \\ -\nabla_y f(z) \end{bmatrix}$$

then, GDA as (discrete-time) dynamical System

$$\boldsymbol{z_{k+1} = w(z_k) = z_k - \eta\Lambda_\tau F(z_k)}$$

Fixed Point Iterations & Dynamical Systems

Gradient Descent Ascent (GDA):

$$x_{k+1} = x_k - \frac{\eta}{\tau}\nabla_x f(x_k, y_k)$$
$$y_{k+1} = y_k + \eta\nabla_y f(x_k, y_k)$$

(step sizes τ, η)

Discrete-time dynamical system:

$$z_{k+1} = w(z_k)$$

If we define

$$z = \begin{bmatrix} x \\ y \end{bmatrix}, \; \Lambda_\tau = \begin{bmatrix} \frac{1}{\tau}I & 0 \\ 0 & I \end{bmatrix}$$
$$F(z) = \begin{bmatrix} \nabla_x f(z) \\ -\nabla_y f(z) \end{bmatrix}$$

then, GDA as (discrete-time) dynamical System

$$\boldsymbol{z_{k+1} = w(z_k) = z_k - \eta\Lambda_\tau F(z_k)}$$

Here,

- an **equilibrium** of the dynamical system (i.e., $w(z^e) = z^e$) satisfies $F(z^e) = 0$
- a **fixed point** of GDA satisfies $F(z_k) = 0$
- a **stationary/critical point** of f satisfies $F(z) = 0$

Fixed Point Iterations & Dynamical Systems

Gradient Descent Ascent (GDA):

$$x_{k+1} = x_k - \frac{\eta}{\tau}\nabla_x f(x_k, y_k)$$
$$y_{k+1} = y_k + \eta\nabla_y f(x_k, y_k)$$

(step sizes τ, η)

Discrete-time dynamical system:

$$z_{k+1} = w(z_k)$$

If we define

$$z = \begin{bmatrix} x \\ y \end{bmatrix}, \ \Lambda_\tau = \begin{bmatrix} \frac{1}{\tau}I & 0 \\ 0 & I \end{bmatrix}$$
$$F(z) = \begin{bmatrix} \nabla_x f(z) \\ -\nabla_y f(z) \end{bmatrix}$$

then, GDA as (discrete-time) dynamical System

$$\boldsymbol{z_{k+1} = w(z_k) = z_k - \eta\Lambda_\tau F(z_k)}$$

Here,

- an **equilibrium** of the dynamical system (i.e., $w(z^e) = z^e$) satisfies $F(z^e) = 0$
- a **fixed point** of GDA satisfies $F(z_k) = 0$
- a **stationary/critical point** of f satisfies $F(z) = 0$

But, what about

- second-order optimality?
- convergence (stability/instability)?
- connections to continuous-time dynamics

$$\dot{z}(t) \approx \frac{z(t+\eta) - z(t)}{\eta} = -\Lambda_\tau F(z(t))$$

Optimality Conditions and Definitions

$$\min_x \max_y f(x,y), \qquad z = (x,y)$$

Optimality Conditions and Definitions

$$\min_x \max_y f(x,y), \qquad z = (x,y)$$

Definition (Lipschitz continuity)

f has globally Lipschitz continuous gradients with Lipschitz constant $L_1(f) > 0$ if

$$\|\nabla f(z_1) - \nabla f(z_2)\| \leq L_1(f)\|z_1 - z_2\| \quad \forall z_1, z_2$$

Optimality Conditions and Definitions

$$\min_x \max_y f(x, y), \qquad z = (x, y)$$

Definition (Lipschitz continuity)

f has globally Lipschitz continuous gradients with Lipschitz constant $L_1(f) > 0$ if

$$\|\nabla f(z_1) - \nabla f(z_2)\| \le L_1(f)\|z_1 - z_2\| \quad \forall z_1, z_2$$

Definition (Stationary (Critical) Points)

Stationary point: $\nabla f(z_0) = 0$

Definition (ϵ-Stationary Points)

ϵ-stationary point: $\|\nabla f(z_0)\| \le \epsilon$ for $z_0 = (x_0, y_0)$ and $\epsilon > 0$.

Optimality Conditions and Definitions

$$\min_x \max_y f(x,y), \qquad z=(x,y) \qquad \text{(Assumption: Finding all stationary points is hard.)}$$

Definition (Lipschitz continuity)

f has globally Lipschitz continuous gradients with Lipschitz constant $L_1(f) > 0$ if

$$\|\nabla f(z_1) - \nabla f(z_2)\| \leq L_1(f)\|z_1 - z_2\| \quad \forall z_1, z_2$$

Definition (Stationary (Critical) Points)

Stationary point: $\nabla f(z_0) = 0$

Definition (ϵ-Stationary Points)

ϵ-stationary point: $\|\nabla f(z_0)\| \leq \epsilon$ for $z_0 = (x_0, y_0)$ and $\epsilon > 0$.

Optimality Conditions and Definitions

$$\min_x \max_y f(x,y), \qquad z = (x,y) \qquad \text{(Assumption: Finding all stationary points is hard.)}$$

Definition (Lipschitz continuity)

f has globally Lipschitz continuous gradients with Lipschitz constant $L_1(f) > 0$ if

$$\|\nabla f(z_1) - \nabla f(z_2)\| \leq L_1(f)\|z_1 - z_2\| \quad \forall z_1, z_2$$

Definition (Stationary (Critical) Points)

Stationary point: $\nabla f(z_0) = 0$

Definition (ϵ-Stationary Points)

ϵ-stationary point: $\|\nabla f(z_0)\| \leq \epsilon$ for $z_0 = (x_0, y_0)$ and $\epsilon > 0$.

Definition (Saddle Points/Nash equilibria)

$z^* = (x^*, y^*)$ is a (local) saddle point of f if:

$$f(x^*, y) \leq f(x^*, y^*) \leq f(x, y^*) \quad \forall (x,y) \in U$$

Proposition (Second order optimality)

Any saddle point (x^*, y^*) is a critical point of f and satisfies

$$\nabla^2_{xx} f(x^*, y^*) \succeq 0 \quad \textit{and} \quad \nabla^2_{yy} f(x^*, y^*) \preceq 0.$$

Second-Order Optimality Analysis: A Dynamical Systems Perspective

Consider $z^+ = w(z)$, $\quad z^e = w(z^e)$, $\quad w(\cdot)$ is a C^1-function

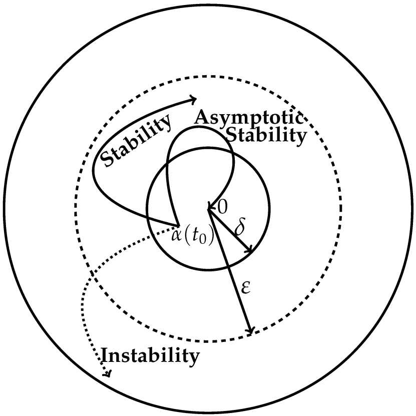
Definition (Stability)

The equilibrium z^e is

- locally **stable** if, $\forall\, \varepsilon > 0\ \exists\, \delta > 0$, such that $z_0 \in \mathcal{B}_\delta(z^e) \implies z_k \in \mathcal{B}_\varepsilon(z^e)\ \forall\, k \geq 0$
- **unstable** if it is not stable
- locally **asymptotically stable** if it is stable and $\exists\, \delta > 0$ such that $\lim_{k\to\infty} z_k = z^e\ \forall\, z_0 \in \mathcal{B}_\delta(z^e)$.

Theorem

- *If* $\rho(\frac{\partial w}{\partial z}(z^e)) < 1$, *then* z^e *is locally asymp. stable.*
- *If* $\rho(\frac{\partial w}{\partial z}(z^e)) > 1$, *then* z^e *is unstable.*



Second-Order Optimality Analysis: A Dynamical Systems Perspective

Consider $z^+ = w(z), \quad z^e = w(z^e), \quad w(\cdot)$ is a C^1-function

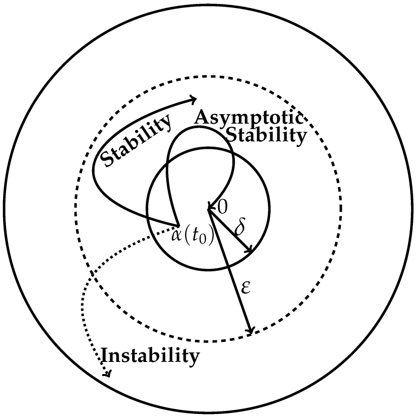
Definition (Stability)

The equilibrium z^e is

- locally **stable** if, $\forall\ \varepsilon > 0\ \exists\ \delta > 0$, such that $z_0 \in \mathcal{B}_\delta(z^e) \implies z_k \in \mathcal{B}_\varepsilon(z^e)\ \forall\ k \geq 0$
- **unstable** if it is not stable
- locally **asymptotically stable** if it is stable and $\exists\ \delta > 0$ such that $\lim_{k\to\infty} z_k = z^e\ \forall\ z_0 \in \mathcal{B}_\delta(z^e)$.

Theorem

- *If* $\rho(\frac{\partial w}{\partial z}(z^e)) < 1$, *then* z^e *is locally asymp. stable.*
- *If* $\rho(\frac{\partial w}{\partial z}(z^e)) > 1$, *then* z^e *is unstable.*



Is there a connection between stability and saddle points/Nash equilibria?

Limitations of GDA

$$z_{k+1} = w(z_k) = z_k - \eta \Lambda_\tau F(z_k), \qquad z = \begin{bmatrix} x \\ y \end{bmatrix}, \ \Lambda_\tau = \begin{bmatrix} \frac{1}{\tau} I & 0 \\ 0 & I \end{bmatrix}, \ F(z) = \begin{bmatrix} \nabla_x f(z) \\ -\nabla_y f(z) \end{bmatrix}$$

Limitations of GDA

$$z_{k+1} = w(z_k) = z_k - \eta\Lambda_\tau F(z_k), \qquad z = \begin{bmatrix} x \\ y \end{bmatrix},\ \Lambda_\tau = \begin{bmatrix} \frac{1}{\tau}I & 0 \\ 0 & I \end{bmatrix},\ F(z) = \begin{bmatrix} \nabla_x f(z) \\ -\nabla_y f(z) \end{bmatrix}$$

Jacobian GDA dynamics:

$$\frac{\partial w}{\partial z}(z_e) = I + \eta\Lambda_\tau H(z_e), \qquad H(z) = \begin{bmatrix} -I & 0 \\ 0 & I \end{bmatrix} \nabla^2 f(z)$$

Limitations of GDA

$$z_{k+1} = w(z_k) = z_k - \eta \Lambda_\tau F(z_k), \qquad z = \begin{bmatrix} x \\ y \end{bmatrix}, \; \Lambda_\tau = \begin{bmatrix} \frac{1}{\tau} I & 0 \\ 0 & I \end{bmatrix}, \; F(z) = \begin{bmatrix} \nabla_x f(z) \\ -\nabla_y f(z) \end{bmatrix}$$

Jacobian GDA dynamics:

$$\frac{\partial w}{\partial z}(z_e) = I + \eta \Lambda_\tau H(z_e), \qquad H(z) = \begin{bmatrix} -I & 0 \\ 0 & I \end{bmatrix} \nabla^2 f(z)$$

GDA diverges for any choice of step sizes, when H has imaginary eigenvalues at the saddle point as $\rho(\frac{\partial w}{\partial z}(z_e)) > 1$.

Limitations of GDA

$$z_{k+1} = w(z_k) = z_k - \eta \Lambda_\tau F(z_k), \qquad z = \begin{bmatrix} x \\ y \end{bmatrix}, \ \Lambda_\tau = \begin{bmatrix} \frac{1}{\tau} I & 0 \\ 0 & I \end{bmatrix}, \ F(z) = \begin{bmatrix} \nabla_x f(z) \\ -\nabla_y f(z) \end{bmatrix}$$

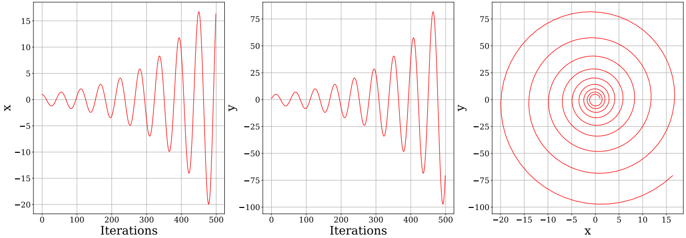
Jacobian GDA dynamics:

$$\frac{\partial w}{\partial z}(z_e) = I + \eta \Lambda_\tau H(z_e), \qquad H(z) = \begin{bmatrix} -I & 0 \\ 0 & I \end{bmatrix} \nabla^2 f(z)$$

GDA diverges for any choice of step sizes, when H has imaginary eigenvalues at the saddle point as $\rho(\frac{\partial w}{\partial z}(z_e)) > 1$.

Example:

- $f(x, y) = xy$
- Even though $(0, 0)$ is a saddle point, GDA does not converge.



Generalised Extra Gradient Algorithm

Generalised Extra Gradient algorithm:

$$\begin{aligned}\hat{x}_k &= x_k - \alpha_{1x}\nabla_x f(x_k, y_k)\\ \hat{y}_k &= y_k + \alpha_{1y}\nabla_y f(x_k, y_k)\\ x_{k+1} &= x_k - \alpha_{2x}\nabla_x f(\hat{x}_k, \hat{y}_k)\\ y_{k+1} &= y_k + \alpha_{2y}\nabla_y f(\hat{x}_k, \hat{y}_k)\end{aligned}$$

Generalised Extra Gradient Algorithm

Generalised Extra Gradient algorithm:

$$\begin{aligned}
\hat{x}_k &= x_k - \alpha_{1x} \nabla_x f(x_k, y_k) \\
\hat{y}_k &= y_k + \alpha_{1y} \nabla_y f(x_k, y_k) \\
x_{k+1} &= x_k - \alpha_{2x} \nabla_x f(\hat{x}_k, \hat{y}_k) \\
y_{k+1} &= y_k + \alpha_{2y} \nabla_y f(\hat{x}_k, \hat{y}_k)
\end{aligned}$$

Step size simplification

$$\gamma = \frac{\alpha_{2x}}{\alpha_{1x}} = \frac{\alpha_{2y}}{\alpha_{1y}}, \qquad \tau = \frac{\alpha_{1y}}{\alpha_{1x}} = \frac{\alpha_{2y}}{\alpha_{2x}}$$

$$\eta = \alpha_{1y}$$

Extra Gradient Algorithms in the Literature:

Algorithm	Parameters	
EG	$\tau = 1$	$\gamma = 1$
τ-EG	$\tau \geq 1$	$\gamma = 1$
EG+	$\tau = 1$	$0 < \gamma \leq 1$
GEG	$\tau > 0$	$\gamma > 0$

Generalised Extra Gradient Algorithm

Generalised Extra Gradient algorithm:

$$\begin{aligned}
\hat{x}_k &= x_k - \alpha_{1x}\nabla_x f(x_k, y_k)\\
\hat{y}_k &= y_k + \alpha_{1y}\nabla_y f(x_k, y_k)\\
x_{k+1} &= x_k - \alpha_{2x}\nabla_x f(\hat{x}_k, \hat{y}_k)\\
y_{k+1} &= y_k + \alpha_{2y}\nabla_y f(\hat{x}_k, \hat{y}_k)
\end{aligned}$$

Step size simplification

$$\gamma = \frac{\alpha_{2x}}{\alpha_{1x}} = \frac{\alpha_{2y}}{\alpha_{1y}}, \qquad \tau = \frac{\alpha_{1y}}{\alpha_{1x}} = \frac{\alpha_{2y}}{\alpha_{2x}}$$

$$\eta = \alpha_{1y}$$

Recall notation:

$$z = \begin{bmatrix} x \\ y \end{bmatrix}, \ \Lambda_\tau = \begin{bmatrix} \frac{1}{\tau}I & 0 \\ 0 & I \end{bmatrix}, \ F(z) = \begin{bmatrix} \nabla_x f(z) \\ -\nabla_y f(z) \end{bmatrix}$$

Extra Gradient Algorithms in the Literature:

Algorithm	Parameters	
EG	$\tau = 1$	$\gamma = 1$
τ-EG	$\tau \geq 1$	$\gamma = 1$
EG+	$\tau = 1$	$0 < \gamma \leq 1$
GEG	$\tau > 0$	$\gamma > 0$

GEG as Dynamical System:

$$z_{k+1} = w(z_k) = z_k - \gamma\eta\Lambda_\tau F(z_k - \eta\Lambda_\tau F(z_k))$$

GEG for $f(x,y) = xy$

GEG as Dynamical System:

$$z_{k+1} = w(z_k) = z_k - \gamma\eta\Lambda_\tau F(z_k - \eta\Lambda_\tau F(z_k))$$

Linearisation of w:

$$\frac{\partial w}{\partial z}(z) = \mathrm{I} + \gamma\eta\Lambda_\tau H(z - \eta\Lambda_\tau F(z))\big(\mathrm{I} + \eta\Lambda_\tau H(z)\big)$$

where

$$H(z) = \begin{bmatrix} -I & 0 \\ 0 & I \end{bmatrix} \nabla^2 f(z)$$

GEG for $f(x,y) = xy$

GEG as Dynamical System:

$$z_{k+1} = w(z_k) = z_k - \gamma\eta\Lambda_\tau F(z_k - \eta\Lambda_\tau F(z_k))$$

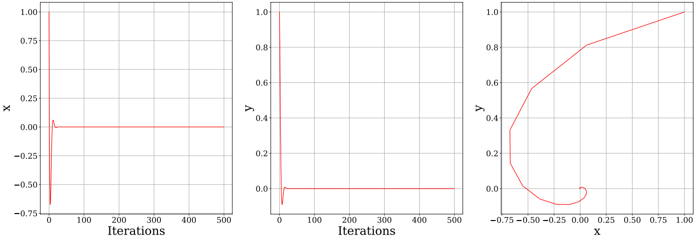
Linearisation of w:

$$\frac{\partial w}{\partial z}(z) = \mathrm{I} + \gamma\eta\Lambda_\tau H(z - \eta\Lambda_\tau F(z))\big(\mathrm{I} + \eta\Lambda_\tau H(z)\big)$$

where

$$H(z) = \begin{bmatrix} -I & 0 \\ 0 & I \end{bmatrix} \nabla^2 f(z)$$

For $f(x,y) = xy$ with $\eta = 0.5,$ $\tau = 0.2$ and $\gamma = 0.25$



GEG Limit Points

GEG: $z_{k+1} = w(z_k) = z_k - \gamma\eta\Lambda_\tau F(z_k - \eta\Lambda_\tau F(z_k))$

Theorem

- *Let f be C^2*
- *Let ∇f be globally Lipschitz with $L > 0$*
- *Let $\eta \in (0, \frac{c}{L})$ for $c > 0$.*

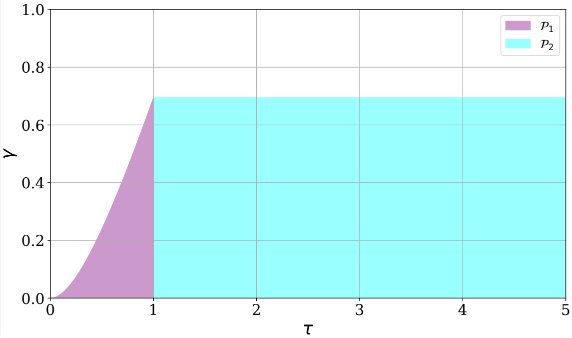
If $(\tau, \gamma) \in \mathcal{P}_1 \cup \mathcal{P}_2$, where

$$\mathcal{P}_1 = \{(\tau,\gamma) | 0 < \tau \leq 1,\ 0 < \gamma \leq \frac{\tau^2}{c\tau + c^2}\}$$
$$\mathcal{P}_2 = \{(\tau,\gamma) | \tau \geq 1,\ 0 < \gamma \leq \frac{1}{c + c^2}\},$$

then

- *for any unstable equilibrium z^* of GEG, $\mathcal{R}(z^*)$ is of measure zero.*

Visualisation of $\mathcal{P}_1 \cup \mathcal{P}_2$ for $c = 0.8$



Region of attraction: $\mathcal{R}(z^*) = \left\{ z_0 \,\middle|\, \lim_{k\to\infty} z_k = z^*,\ z_{k+1} = w(z_k) \right\}$

GEG Limit Points

GEG: $$z_{k+1} = w(z_k) = z_k - \gamma\eta\Lambda_\tau F(z_k - \eta\Lambda_\tau F(z_k))$$

Theorem

Let f be C^2, $\nabla^2 f(z^*)$ be invertible and let $\tau > 0$, $\eta \in (0, \frac{\min\{1,\tau\}}{L})$.

- *If* $[\sigma(\Lambda_\tau H(z^*)) \subset \mathbb{R}$ *and* $\gamma \in (0, 8)]$ *OR* $\gamma \in (0, 1]$

Then the set of saddle points of f is a subset of the set of locally asymptotically stable equilibria of GEG.

GEG Limit Points

GEG: $z_{k+1} = w(z_k) = z_k - \gamma\eta\Lambda_\tau F(z_k - \eta\Lambda_\tau F(z_k))$

Theorem

Let f be C^2, $\nabla^2 f(z^*)$ be invertible and let $\tau > 0$, $\eta \in (0, \frac{\min\{1,\tau\}}{L})$.

- *If $[\sigma(\Lambda_\tau H(z^*)) \subset \mathbb{R}$ and $\gamma \in (0,8)]$ OR $\gamma \in (0,1]$*

Then the set of saddle points of f is a subset of the set of locally asymptotically stable equilibria of GEG.

Remark

There are functions f and asymptotically stable equilibria z^* of GEG, which do not characterise saddle points of f.

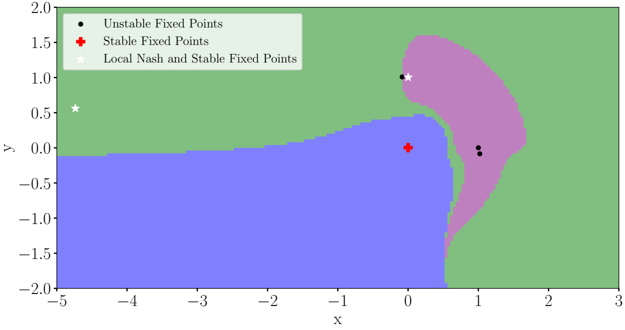
Numerical Example (Function With Several Equilibria)

$$f(x, y) = f_1(x, y)(x-1)^2(y-1)^2 + f_2(x, y)x^2y^2,$$

$$f_1(x, y) = -0.25x^2 - 0.5y^2 + 0.6xy$$

$$f_2(x, y) = 0.5x^2 + 0.5y^2 + 4xy$$

Equilibria	GEG-stable	Saddle point
$(0, 0)$	YES	NO
$(0, 1)$	YES	YES
$(1, 0)$	NO	NO
$(-4.73, 0.56)$	YES	YES
$(1.02, -0.09)$	NO	NO
$(0.73, -5.40)$	NO	NO
$(-0.09, 1.01)$	NO	NO
$(38.40, -1.49)$	YES	YES



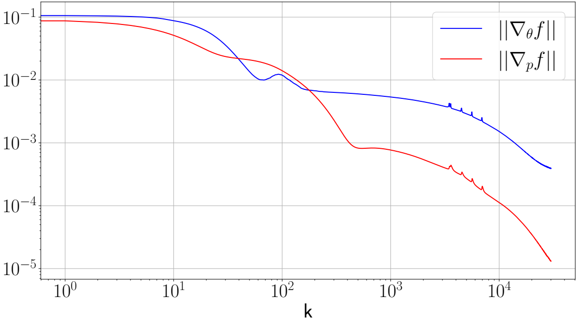
Classification of critical points:

- (Black) unstable equilibria
- (Red/White) asymp. stable equilibria
- (White) asymp. stable equilibria & saddle points

Empirical Risk Minimisation for a Binary Classification Using NN

$$\min_{\theta} \max_{p} \; -\sum_{i=1}^{m} p_i \left[y_i \log(\hat{y}(X_i;\theta)) + (1-y_i)\log(1-\hat{y}(X_i;\theta)) \right] - \psi \sum_{i=1}^{m} \left(p_i - \frac{1}{m}\right)^2.$$

$\hat{y}(X;\theta)$ is generated by a single-hidden-layer network (size 50, LeakyReLU) with $n = 1601$, $m = 455$.



Starting from random initialisations, where

- $\sigma(\nabla^2_{\theta\theta} f(z_0)) \subset (-10, 10)$,

the iterates converge to limit points where

- $\nabla f(z^*) \approx 0$,
- $\nabla^2_{pp} f(z^*) \prec 0$,
- $\sigma(\nabla^2_{\theta\theta} f(z^*)) \subset (-10^{-5}, 10^2)$.

Intermediate summary

So far, we have ...

- considered iterative fixed point methods
- written the fixed point methods as discrete-time dynamical systems

$$z_{k+1} = w(z_k)$$

- analysed stability properties of equilibria/fixed points z^* based on eigenvalue cond.:

$$\rho\left(\frac{\partial w}{\partial z}(z^*)\right) < 1 \qquad (\text{or } > 1)$$

Intermediate summary

So far, we have ...

- considered iterative fixed point methods
- written the fixed point methods as discrete-time dynamical systems

$$z_{k+1} = w(z_k)$$

- analysed stability properties of equilibria/fixed points z^* based on eigenvalue cond.:

$$\rho\left(\frac{\partial w}{\partial z}(z^*)\right) < 1 \qquad (\text{or } \; > 1)$$

But ...

- the eigenvalue condition is limited.
- can we also identify continuous-time dynamical systems $\dot{z} = \bar{w}(z)$ for fixed point methods (and analyse stability properties of their equilibria)?
  - Can we use the results to derive novel iterative algorithms?

Stability revisited

Consider $z^+ = w(z), \quad z^e = w(z^e),$

Definition (Stability)

The equilibrium z^e is

- locally **stable** if, $\forall\ \varepsilon > 0\ \exists\ \delta > 0$, s.t. $z_0 \in \mathcal{B}_\delta(z^e) \implies z_k \in \mathcal{B}_\varepsilon(z^e)\ \forall\ k \geq 0$
- **unstable** if it is not stable
- locally **asymp. stable** if it is stable and $\exists\ \delta > 0$ s.t. $\lim_{k\to\infty} \|z_k\|_{z^e} = 0\ \forall\ z_0 \in \mathcal{B}_\delta(z^e)$
- locally **exp. stable** if it is stable and $\exists\ M, \delta > 0,\ \rho < 1$, s.t. $\|z_k\|_{z^e} \leq M\|z_0\|_{z^e}\rho^k\ \forall k\ \forall\ z_0 \in \mathcal{B}_\delta(z^e)$

Stability revisited

Consider $z^+ = w(z), \quad z^e = w(z^e),$

Definition (Stability)

The equilibrium z^e is

- locally **stable** if, $\forall\ \varepsilon > 0\ \exists\ \delta > 0$, s.t. $z_0 \in \mathcal{B}_\delta(z^e) \implies z_k \in \mathcal{B}_\varepsilon(z^e)\ \forall\ k \geq 0$
- **unstable** if it is not stable
- locally **asymp. stable** if it is stable and $\exists\ \delta > 0$ s.t. $\lim_{k\to\infty} \|z_k\|_{z^e} = 0\ \forall\ z_0 \in \mathcal{B}_\delta(z^e)$
- locally **exp. stable** if it is stable and $\exists\ M, \delta > 0,\ \rho < 1$, s.t. $\|z_k\|_{z^e} \leq M\|z_0\|_{z^e}\rho^k\ \forall k\ \forall\ z_0 \in \mathcal{B}_\delta(z^e)$

Consider $\dot{z} = \bar{w}(z), \quad 0 = \bar{w}(z^e),$

Definition (Stability)

The equilibrium z^e is

- locally **stable** if, $\forall\ \varepsilon > 0\ \exists\ \delta > 0$, s.t. $z_0 \in \mathcal{B}_\delta(z^e) \implies z(t) \in \mathcal{B}_\varepsilon(z^e)\ \forall\ t \in \mathbb{R}_{\geq 0}$
- **unstable** if it is not stable
- locally **asymp. stable** if it is stable and $\exists\ \delta > 0$ s.t. $\lim_{t\to\infty} \|z(t)\|_{z^e} = 0\ \forall\ z_0 \in \mathcal{B}_\delta(z^e)$
- locally **exp. stable** if it is stable and $\exists\ M, \delta, \lambda > 0$, s.t. $\|z(t)\|_{z^e} \leq M\|z_0\|_{z^e}e^{-\lambda t}\ \forall t,\ \forall\ z_0 \in \mathcal{B}_\delta(z^e)$

Stability revisited

Consider $z^+ = w(z), \quad z^e = w(z^e),$

Definition (Stability)

The equilibrium z^e is

- locally **stable** if, $\forall\ \varepsilon > 0\ \exists\ \delta > 0$, s.t. $z_0 \in \mathcal{B}_\delta(z^e) \implies z_k \in \mathcal{B}_\varepsilon(z^e)\ \forall\ k \geq 0$
- **unstable** if it is not stable
- locally **asymp. stable** if it is stable and $\exists\ \delta > 0$ s.t. $\lim_{k\to\infty} \|z_k\|_{z^e} = 0\ \forall\ z_0 \in \mathcal{B}_\delta(z^e)$
- locally **exp. stable** if it is stable and $\exists\ M, \delta > 0,\ \rho < 1$, s.t. $\|z_k\|_{z^e} \leq M\|z_0\|_{z^e}\rho^k\ \forall k\ \forall\ z_0 \in \mathcal{B}_\delta(z^e)$

Consider $\dot{z} = \bar{w}(z), \quad 0 = \bar{w}(z^e),$

Definition (Stability)

The equilibrium z^e is

- locally **stable** if, $\forall\ \varepsilon > 0\ \exists\ \delta > 0$, s.t. $z_0 \in \mathcal{B}_\delta(z^e) \implies z(t) \in \mathcal{B}_\varepsilon(z^e)\ \forall\ t \in \mathbb{R}_{\geq 0}$
- **unstable** if it is not stable
- locally **asymp. stable** if it is stable and $\exists\ \delta > 0$ s.t. $\lim_{t\to\infty} \|z(t)\|_{z^e} = 0\ \forall\ z_0 \in \mathcal{B}_\delta(z^e)$
- locally **exp. stable** if it is stable and $\exists\ M, \delta, \lambda > 0$, s.t. $\|z(t)\|_{z^e} \leq M\|z_0\|_{z^e} e^{-\lambda t}\ \forall t,\ \forall\ z_0 \in \mathcal{B}_\delta(z^e)$

Can be stated in terms of invariant sets: $\|\cdot\|_{\mathcal{A}}, \quad \mathcal{A} \subset \mathbb{R}^d$ closed (and bounded)

Stability revisited (Lyapunov characterizations)

Consider $z^+ = w(z), \quad z^e = w(z^e)$,

Definition (Stability)

The equilibrium z^e is

- locally **exp. stable** if it is stable and $\exists\ M, \delta > 0,\ \rho < 1$, s.t. $\|z_k\|_{z^e} \leq M\|z_0\|_{z^e}\rho^k\ \forall k\ \forall\ z_0 \in \mathcal{B}_\delta(z^e)$

Theorem (Lyapunov characterisation)

Equilibrium z^e is exp. stable iff there exist $\lambda_1, \lambda_2, \delta > 0$, $c \in (0,1)$ and $V : \mathcal{B}_\delta(z^e) \to \mathbb{R}^d$ s.t.

$$\begin{aligned} \lambda_1\|z\|^2_{z_e} \leq V(z) \leq \lambda_2\|z\|^2_{z_e} \\ V(z^+) - V(z) \leq -cV(z) \quad \forall z \in \mathcal{B}_\delta(z^e) \end{aligned}$$

Stability revisited (Lyapunov characterizations)

Consider $z^+ = w(z), \quad z^e = w(z^e),$

Definition (Stability)

The equilibrium z^e is

- locally **exp. stable** if it is stable and $\exists\ M, \delta > 0,\ \rho < 1$, s.t. $\|z_k\|_{z^e} \leq M\|z_0\|_{z^e}\rho^k\ \forall k\ \forall\ z_0 \in \mathcal{B}_\delta(z^e)$

Theorem (Lyapunov characterisation)

Equilibrium z^e is exp. stable iff there exist $\lambda_1, \lambda_2, \delta > 0,\ c \in (0,1)$ and $V : \mathcal{B}_\delta(z^e) \to \mathbb{R}^d$ s.t.

$$\lambda_1\|z\|^2_{z_e} \leq V(z) \leq \lambda_2\|z\|^2_{z_e}$$
$$V(z^+) - V(z) \leq -cV(z) \quad \forall z \in \mathcal{B}_\delta(z^e)$$

Consider $\dot{z} = \bar{w}(z), \quad 0 = w(z^e),$

Definition (Stability)

The equilibrium z^e is

- locally **exp. stable** if it is stable and $\exists\ M, \delta, \lambda > 0$, s.t. $\|z(t)\|_{z^e} \leq M\|z_0\|_{z^e}e^{-\lambda t}\ \forall t,\ \forall\ z_0 \in \mathcal{B}_\delta(z^e)$

Theorem (Lyapunov characterisation)

Equilibrium z^e is exp. stable iff there exist $\lambda_1, \lambda_2, \delta > 0,\ c > 0$ and $V : \mathcal{B}_\delta(z^e) \to \mathbb{R}^d$ (C^1) s.t.

$$\lambda_1\|z\|^2_{z_e} \leq V(z) \leq \lambda_2\|z\|^2_{z_e}$$
$$\langle \nabla V(z), w(z)\rangle \leq -cV(z) \quad \forall z \in \mathcal{B}_\delta(z^e)$$

Stability Link Between Discrete and Continuous Dynamics

Consider $\dot{x} = \bar{w}_s(z)$ and $z^+ = w_s(z)$ for $s > 0$ and common equilibrium z^e.

Assumption (Consistency of order r)

For $r \geq 2$, suppose that for any z_0 we have

$$\|z(s; z_0) - z^+\| = \|z(s; z_0) - w_s(z_0)\| \leq O(s^r).$$

Stability Link Between Discrete and Continuous Dynamics

Consider $\dot{x} = \bar{w}_s(z)$ and $z^+ = w_s(z)$ for $s > 0$ and common equilibrium z^e.

Assumption (Consistency of order r)

For $r \geq 2$, suppose that for any z_0 we have

$$\|z(s;z_0) - z^+\| = \|z(s;z_0) - w_s(z_0)\| \leq O(s^r).$$

Example (Explicit Euler discretisation)

Consider $\dot{z} = f(z)$ (with $f(0) = 0$)) and $z^+ = z + sf(z)$.
Taylor approximation: $z(t+s) = z(t) + s\dot{z}(t) + s^2\ddot{z}(t+\delta) = z(t) + sf(z(t)) + O(s^2)$

Stability Link Between Discrete and Continuous Dynamics

Consider $\dot{x} = \bar{w}_s(z)$ and $z^+ = w_s(z)$ for $s > 0$ and common equilibrium z^e.

Assumption (Consistency of order r)

For $r \geq 2$, suppose that for any z_0 we have

$$\|z(s; z_0) - z^+\| = \|z(s; z_0) - w_s(z_0)\| \leq O(s^r).$$

Theorem

- *If z^e is locally exp./asympt. stable wrt $\bar{w}_s(\cdot)$,*
- *Then $\exists\ s^*$ such that $\forall\ s \in (0, s^*]$, z^e is a locally exp./asymp. stable wrt. $w_s(\cdot)$*

Example (Explicit Euler discretisation)

Consider $\dot{z} = f(z)$ (with $f(0) = 0$)) and $z^+ = z + sf(z)$.
Taylor approximation: $z(t + s) = z(t) + s\dot{z}(t) + s^2\ddot{z}(t + \delta) = z(t) + sf(z(t)) + O(s^2)$

From discrete-time to continuous-time

Definition ($O(s^r)$-resolution ODE)

For $r \in \mathbb{N}$, consider $\dot{z} = \bar{w}_s^r(z) = f_0(z) + s f_1(z) + \cdots + s^r f_r(z)$, $\quad f_i : \mathbb{R}^d \to \mathbb{R}^d, C^1$-fcns..
Then $\bar{w}_s^r(z)$ is called $O(s^r)$-resolution ODE of $z^+ = w_s(z)$, if it satisfies

$$\|z(s; z_0) - w_s(z_0)\| = O(s^r) \qquad \forall\ z_0 \in \mathbb{R}^d.$$

From discrete-time to continuous-time

Definition ($O(s^r)$-resolution ODE)

For $r \in \mathbb{N}$, consider $\dot{z} = \bar{w}_s^r(z) = f_0(z) + s f_1(z) + \cdots + s^r f_r(z)$, $\quad f_i : \mathbb{R}^d \to \mathbb{R}^d, C^1$-fcns..
Then $\bar{w}_s^r(z)$ is called $O(s^r)$-resolution ODE of $z^+ = w_s(z)$, if it satisfies

$$\|z(s; z_0) - w_s(z_0)\| = O(s^r) \qquad \forall\ z_0 \in \mathbb{R}^d.$$

Theorem ($O(s^r)$-resolution ODE construction)

Consider $z^+ = w_s(z)$ with $w_0(z) = z\ \forall\ z \in \mathbb{R}^d$. Then the $O(s^r)$-resolution ODE is unique and coefficient functions of $\bar{w}_s^r(z)$ can be obtained recursively, i.e.,

$$f_i(z) = \tfrac{1}{(i+1)!} \tfrac{\partial^{i+1} w_s(z)}{\partial s^{i+1}}\Big|_{s=0} - \textstyle\sum_{l=2}^{i+1} \tfrac{1}{l!} h_{l,i+1-l}(z), \quad \forall\ i = 0, 1, \ldots, r,$$

where $h_{0,0}(z) = z$, $h_{1,i}(z) = f_i(z)$ for $i \in \{0, \ldots, r\}$ and

$$h_{j+1,i}(z) = \textstyle\sum_{l=0}^{i} \nabla h_{j,l}(z) h_{1,i-l}(z), \quad h_{0,i}(z) = 0 \quad \text{for } i, j \in \{1, \ldots, r\}.$$

Lu, H. An $O(s^r)$-resolution ODE framework for understanding discrete-time algorithms and applications to the linear convergence of minimax problems. Math. Program. 194, 1061–1112 (2022).

Generalised Extra Gradient Algorithm (revisited)

Generalised Extra Gradient algorithm:

$$\begin{aligned}
\hat{x}_k &= x_k - \alpha_{1x}\nabla_x f(x_k, y_k) \\
\hat{y}_k &= y_k + \alpha_{1y}\nabla_y f(x_k, y_k) \\
x_{k+1} &= x_k - \alpha_{2x}\nabla_x f(\hat{x}_k, \hat{y}_k) \\
y_{k+1} &= y_k + \alpha_{2y}\nabla_y f(\hat{x}_k, \hat{y}_k)
\end{aligned}$$

Recall notation:

$$z = \begin{bmatrix} x \\ y \end{bmatrix}, \ \Lambda_\tau = \begin{bmatrix} \frac{1}{\tau}I & 0 \\ 0 & I \end{bmatrix}, \ F(z) = \begin{bmatrix} \nabla_x f(z) \\ -\nabla_y f(z) \end{bmatrix}$$

GEG as Dynamical System:

$$z_{k+1} = w(z_k) = z_k - \gamma\eta\Lambda_\tau F(z_k - \eta\Lambda_\tau F(z_k))$$

GEG $O(1)$- and $O(s)$-resolution ODEs:

$$\dot{z} = -\gamma\Lambda_\tau F(z)$$

$$\begin{aligned}
\dot{z} &= -\gamma\Lambda_\tau F(z) + s(\gamma\Lambda_\tau\nabla F(z)\Lambda_\tau F(z) - \tfrac{\gamma^2}{2}\Lambda_\tau\nabla F(z)\Lambda_\tau F(z)) \\
&= (-I + (1 - \tfrac{\gamma}{2})s\Lambda_\tau\nabla F(z))\gamma\Lambda_\tau F(z)
\end{aligned}$$

GEG and $O(s^r)$-resolution ODEs (results)

Theorem

Let z^* be a saddle point of f, let f satisfy technical assumptions and define

$$M = \begin{cases} \frac{|\max_{\lambda\in\sigma(\Lambda_\tau H(z^*))}\{\Re(\lambda)\}|}{\max\{L^2, \frac{L^2}{\tau^2}\}} & \text{if} \quad \max_{\lambda\in\sigma(\Lambda_\tau H(z^*))}\Re(\lambda) < 0 \\ \infty & \text{if} \quad \max_{\lambda\in\sigma(\Lambda_\tau H(z^*))}\Re(\lambda) = 0 \end{cases}$$

(i) *Consider the $O(1)$-resolution ODE of GEG and let $\tau > 0$ and $\gamma > 0$.*
Saddle points of f are exp. stable equilibria of GEG if $\Re(\lambda) \neq 0$ for $\lambda \in \sigma(\Lambda_\tau H(z^*))$.

(ii) *Consider the $O(s)$-resolution ODE of GEG and let $\tau > 0$.*
Saddle points of f are exp. stable equilibria of GEG if $0 < \gamma < 2$ and
$0 < s < \min\left\{M, \min\{L^{-1}, \frac{\tau}{L}\}\right\}$.

GEG and $O(s^r)$-resolution ODEs (results)

Theorem

Let z^* be a saddle point of f, let f satisfy technical assumptions and define

$$M = \begin{cases} \frac{|\max_{\lambda \in \sigma(\Lambda_\tau H(z^*))} \{\Re(\lambda)\}|}{\max\{L^2, \frac{L^2}{\tau^2}\}} & \text{if} \quad \max_{\lambda \in \sigma(\Lambda_\tau H(z^*)))} \Re(\lambda) < 0 \\ \infty & \text{if} \quad \max_{\lambda \in \sigma(\Lambda_\tau H(z^*)))} \Re(\lambda) = 0 \end{cases}$$

(i) *Consider the $O(1)$-resolution ODE of GEG and let $\tau > 0$ and $\gamma > 0$. Saddle points of f are exp. stable equilibria of GEG if $\Re(\lambda) \neq 0$ for $\lambda \in \sigma(\Lambda_\tau H(z^*))$.*

(ii) *Consider the $O(s)$-resolution ODE of GEG and let $\tau > 0$. Saddle points of f are exp. stable equilibria of GEG if $0 < \gamma < 2$ and $0 < s < \min\left\{M, \min\{L^{-1}, \frac{\tau}{L}\}\right\}$.*

Interpretation:

- We can use (arbitrary) ODE-solvers to find saddle points
- Can we use this property to come up with novel fixed-point methods?

Conclusions

- Main motivation: Min-Max problems

$$\min_x \max_y f(x,y), \qquad z=(x,y),$$

  i.e., we want to find saddle points of $f(z)$.
- Saddle points satisfy $\nabla f(z)=0$ (but finding all critical points is hard).
- We study fixed point methods by investigating stability properties of equilibria of discrete-time and continuous-time dynamical systems

$$z^+ = w_s(z) \qquad \text{and} \qquad \dot{z} = \bar{w}_s(z)$$

- For GEG, under appropriate step size selections (s), we show that saddle points are exp. stable equilibria (i.e., we don't need to find all critical points).
- Can this analysis be used for algorithm design?
- A similar analysis can be performed for other fixed point methods
  - Two-Timescale Gradient Descent Ascent; (Two-Timescale) Proximal Point Method; (Regularised) Damped Newton; ...