L2. SC-SC: GDA, by monotone field

- monotone field $\quad F: \mathbb{R}^d \to \mathbb{R}^d$ $\quad \mathcal{X}$ convex in $\mathbb{R}^d$

def $\quad F(x)$ is called monotone on $\underline{\mathbb{R}^d}$ if

$$\langle F(\tilde{x}) - F(x), \tilde{x} - x \rangle \geq 0, \ \forall x, \tilde{x} \in \mathbb{R}^d$$

F is α-strongly monotone if

$$\langle F(\tilde{x}) - F(x), \tilde{x} - x \rangle \geq \alpha \|\tilde{x} - x\|^2, \ \forall \tilde{x}, x \in \mathbb{R}^d.$$

- follow monotone field imply convergence

$\underline{\text{lemma}}$ if F is α-strongly monotone, then zero pt of F is unique.

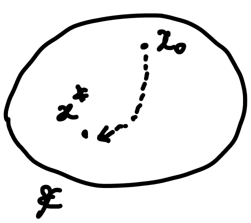
pf. $F(x_1) = F(x_2) = 0$, then

$$\langle F(x_1) - \cancel{F(x_2)}^{\,0}, x_1 - x_2 \rangle \geq \alpha \|x_1 - x_2\|^2 \Rightarrow x_1 = x_2$$

#

Rk F is cont, $\mathcal{X}$ is compact convex, then $\exists!\, x^*$.

- how to converge to x^*?



$$x_{k+1} = x_k - s F(x_k)$$

$$F(x^*) = 0$$

$$\|x_{k+1} - x^*\|^2 = \|x_k - x^* - s F(x_k)\|^2 \quad \leq ?$$

$$= \|x_k - x^*\|^2 - 2s \underbrace{\langle F(x_k), x_k - x^* \rangle}_{\langle F(x_k) - F(x^*), x_k - x^* \rangle \geq \alpha \|x_k - x^*\|^2} + s^2 \|F(x_k)\|^2$$

$$\|F(x_k) - \cancel{F(x^*)}^{\,0}\| \leq \beta \|x_k - x^*\| \quad \text{if } \underline{F \text{ is } \beta\text{-Lipschitz}}$$

$$\beta \geq \alpha > 0$$

then

$$\|x_{k+1} - x^*\|^2 \le \|x_k - x^*\|^2 - 2s\alpha\|x_k - x^*\|^2 + s^2\beta^2\|x_k - x^*\|^2$$

$$= \underbrace{(1 - 2s\alpha + s^2\beta^2)}_{<1 \text{ if } 0 < s < \frac{2\alpha}{\beta^2} \left(\le \frac{2}{\beta}\right)} \|x_k - x^*\|^2$$

this gives $\|x_k - x^*\| \le \rho^k \|x_0 - x^*\|$, $\rho < 1$.

"exponential convergence" i.e.
(linear)

$\|x_k - x^*\| \le \varepsilon$ in $O(\log \frac{1}{\varepsilon})$ steps.

Rm. Follow a strongly monotone field converges to its unique vanishing pt exponentially fast.

Eg. $\min_x f(x)$, f α-(strongly) convex

$\Rightarrow \nabla f$ α-monotone

GD (Gradient Descent)

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

$\Rightarrow$ For strongly convex minimization, GD converges exponentially fast (need $s < \frac{2}{\beta}$)

- difference btw convex v.s. convex-concave

consider $f: \mathbb{R}^2 \to \mathbb{R}$, f is C^2, $\alpha \ge 0$.
$f(x,y) = f(z)$, $z = \begin{bmatrix} x \\ y \end{bmatrix}$

1. (jointly) convex f

fact ∇f is α-monotone $\Leftrightarrow$ $\nabla^2 f(z) \succeq \alpha I$, $\forall z \in \mathbb{R}^2$

$$\nabla^2 f = \begin{bmatrix} \partial^2_{xx} f & \partial^2_{xy} f \\ \partial^2_{yx} f & \partial^2_{yy} f \end{bmatrix}$$ the matrix matter.

pf. "$\Leftarrow$" $\forall z$, $\tilde{z} = z + sh$, $\forall h$, $\|h\| = 1$, $s > 0$,

$$\langle \nabla f(z+sh) - \nabla f(z), sh \rangle \overset{?}{\geq} \alpha \|sh\|^2$$

$$\phi(s) := \langle \nabla f(z+sh), h \rangle ,$$

$$\phi(s) - \phi(0) \overset{?}{\geq} \alpha s \|h\|^2$$

By MVT, $\phi(s) - \phi(0) = s\phi'(\tilde{s})$

$$= s \underbrace{\langle [\nabla^2 f(z+\tilde{s}h)]h, h \rangle}_{\geq \alpha \|h\|^2} \checkmark$$

"$\Rightarrow$" $\forall z$, WTS $\nabla^2 f(z) \succeq \alpha I$.

i.e. $\forall h, \|h\|=1, \langle [\nabla^2 f(z)]h, h \rangle \overset{?}{\geq} \alpha \|h\|^2$

define $\phi(s)$ as above, by F is α-monotone,

$\phi(s) - \phi(0) \geq \alpha s \|h\|^2, \ \forall h, \|h\|=1, \forall s>0.$

$\Rightarrow \phi'(0) \geq \alpha \|h\|^2. \quad \phi'(0) = \langle [\nabla^2 f(z)]h, h \rangle$ #

2. convex-concave f

lemma If $\partial^2_{xx} f(z), -\partial^2_{yy} f(z) \geq \alpha$, then

$$\partial f(z) := \begin{bmatrix} \partial_x f(z) \\ -\partial_y f(z) \end{bmatrix} \text{ is } \alpha\text{-monotone on } \mathbb{R}^2.$$

$$\nabla^2 f = \begin{bmatrix} \underline{\partial^2_{xx} f} & \partial^2_{xy} f \\ \partial^2_{yx} f & \underline{\partial^2_{yy} f} \end{bmatrix}$$ only $\partial_{xx}, \partial_{yy}$ matter!

Cor. $f(x,y) : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}$,

$[\partial^2_{xx} f]_{m\times m} \succeq \alpha I, \ [-\partial^2_{yy} f]_{n \times n} \succeq \alpha I_n,$

then $\partial f(z)$ is α-monotone on $\mathbb{R}^{n+m}$.

pf. $f(\tilde{x},y) \overset{\text{convex}}{\geq} f(x,y) + \partial_x f(x,y)(\tilde{x}-x) + \frac{\alpha}{2}|\tilde{x}-x|^2$

$f(\tilde{x},y) \overset{\text{concave}}{\leq} f(\tilde{x},\tilde{y}) + \partial_y f(\tilde{x},\tilde{y})(y-\tilde{y}) - \frac{\alpha}{2}|\tilde{y}-y|^2$

$\Rightarrow f(\tilde{x},\tilde{y}) - f(x,y) \geq \partial_x f(x,y)(\tilde{x}-x) + \partial_y f(\tilde{x},\tilde{y})(\tilde{y}-y)$

$$+\frac{\alpha}{2}|\tilde{y}-y|^2+\frac{\alpha}{2}|\tilde{x}-x|^2 \quad \cdots ①$$

by symmetry, $(x,y) \leftrightarrow (\tilde{x},\tilde{y})$

$$f(x,y)-f(\tilde{x},\tilde{y}) \geq \partial_x f(\tilde{x},\tilde{y})(x-\tilde{x})+\partial_y f(x,y)(y-\tilde{y})$$
$$+\frac{\alpha}{2}|y-\tilde{y}|^2+\frac{\alpha}{2}|x-\tilde{x}|^2 \quad \cdots ②$$

①, ② $\Rightarrow$

$$\partial_x f(\tilde{x},\tilde{y})(\tilde{x}-x)+\partial_y f(x,y)(\tilde{y}-y)-\frac{\alpha}{2}|y-\tilde{y}|^2-\frac{\alpha}{2}|x-\tilde{x}|^2$$

$$\geq \partial_x f(x,y)(\tilde{x}-x)+\partial_y f(\tilde{x},\tilde{y})(\tilde{y}-y)+\frac{\alpha}{2}|y-\tilde{y}|^2+\frac{\alpha}{2}|x-\tilde{x}|^2$$

i.e.

$$\left\langle \begin{bmatrix} \partial_x f(\tilde{x},\tilde{y})-\partial_x f(x,y) \\ (-\partial_y f)(\tilde{x},\tilde{y})-(-\partial_y f)(x,y) \end{bmatrix}, \begin{bmatrix} \tilde{x}-x \\ \tilde{y}-y \end{bmatrix} \right\rangle \geq \alpha \left\| \begin{bmatrix} \tilde{x}-x \\ \tilde{y}-y \end{bmatrix} \right\|^2$$

i.e. $\langle \partial f(\tilde{z})-\partial f(z), \tilde{z}-z \rangle \geq \alpha \|\tilde{z}-z\|^2$ #

- exponential convergence for SC-SC objective

$$\min_x \max_y L(x,y), \qquad L: \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R},\ C^2$$

a) μ - strongly convex - strongly concave : $\mu > 0$.

$$\nabla^2_{xx} L(x,y) \succeq \mu I_n,\ -\nabla^2_{yy} L(x,y) \succeq \mu I_m,\ \forall x,y$$

b) l - smooth $\quad l \geq \mu > 0$

$\nabla_x L,\ \nabla_y L$ is l-Lipschitz

<u>GDA</u> (Gradient Descent Ascent) $\quad s>0$

$$\begin{cases} x_{k+1} = x_k - s\nabla_x L(x_k,y_k) \\ y_{k+1} = y_k + s\nabla_y L(x_k,y_k) \end{cases}$$

$$z_k = \begin{bmatrix} x_k \\ y_k \end{bmatrix}, \quad \partial L(z) := \begin{bmatrix} \nabla_x L(z) \\ -\nabla_y L(z) \end{bmatrix}$$

$$z_{k+1} = z_k - s\,\partial L(z_k),$$

a) $\Rightarrow$ ∂L is μ-monotone
b) $\Rightarrow$ ∂L is ℓ-Lip
$\Big\}$ $\Rightarrow$ exp. convergence to the ~~a~~ stationary pt
$\partial L(z^*) = 0$

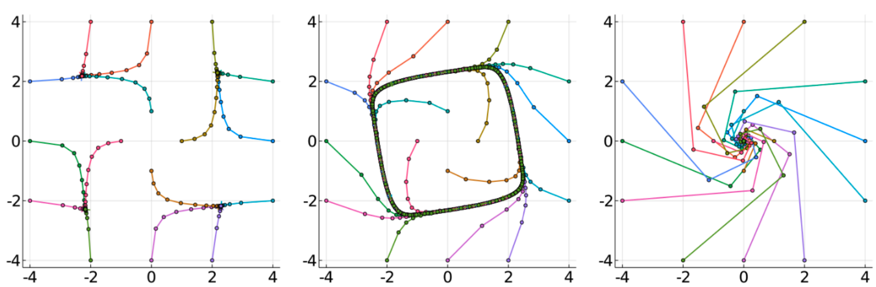
Q. existence of z^*? μ-S.C.-S.C. implies that.

ref lemma B.1 of [GLWM 2023] on PPM

[GLWM 2023] Grimmer, Lu, Worah, Mirrokni (2023). The landscape of the proximal point method for nonconvex–nonconcave minimax optimization. *Mathematical Programming*, *201*(1), 373-407.

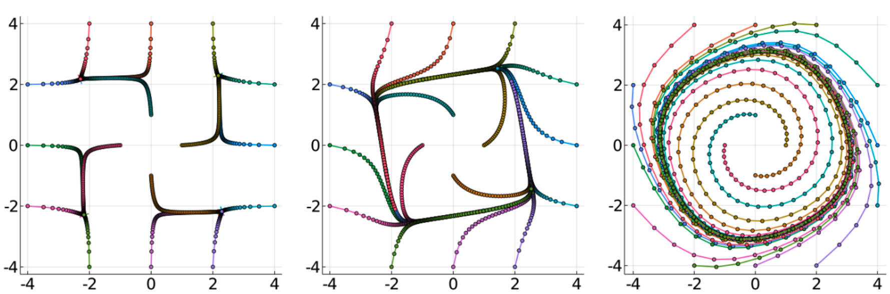
Eg $L(x,y) = f(x) - g(y) + a\,xy$ (bilinear interaction)

effect of a small or large?



(a) $A = 1$ (b) $A = 10$ (c) $A = 100$

Figure 1: Sample paths of PPM from different initial solutions applied to (3) with $f(x) = (x+3)(x+1)(x-1)(x-3)$ and $g(y) = (y+3)(y+1)(y-1)(y-3)$ and different scalars A. As $A \geq 0$ increases, the solution path transitions from having four locally attractive stationary points, to a globally attractive cycle, and finally to a globally attractive stationary point.



(i) GDA w/ A=1 (j) GDA w/ A=10 (k) GDA w/ A=100

This minimax problem is NC-NC type, GDA cycles in a case ($A=100$, strong interaction) when PPM converges better. See L3.