L3. NC-NC: PPM and saddle envelope

$$\min_x \max_y L(x,y), \quad x \in \mathbb{R}^n, \ y \in \mathbb{R}^m.$$

(A1) <u>ℓ-smoothness</u>

L is C^1 on $\mathbb{R}^n \times \mathbb{R}^m$ and $\nabla_x L$, $\nabla_y L$ is ℓ-Lip.

- When L is μ-SC-SC, <u>GDA</u> converges exponentially fast.

<u>Q</u>. what if NC-NC? or SC-NC?
C-C
C-NC

idea: Proximal Point Method (PPM)

$$\begin{bmatrix} x_{k+1} \\ y_{k+1} \end{bmatrix} = \mathrm{prox}_s \begin{bmatrix} x_k \\ y_k \end{bmatrix} \qquad s>0 \text{ step-size}$$

$$= \operatorname*{argmin}_u \max_v L(u,v) + \frac{1}{2s}\|u-x_k\|^2 - \frac{1}{2s}\|v-y_k\|^2.$$

<u>def</u> (saddle envelope)

$$L_s(x,y) = \min_u \max_v L(u,v) + \frac{1}{2s}\|u-x\|^2 - \frac{1}{2s}\|v-y\|^2$$

(A2) <u>WC-WC</u>. L is C^2, and

$\rho>0$ s.t. $\nabla^2_{xx} L \succeq -\rho I$, $-\nabla^2_{yy} L \succeq -\rho I$ on $\mathbb{R}^n \times \mathbb{R}^m$

<u>fact</u> If $\frac{1}{s} > \rho$, then $\operatorname*{argmin}_u$, $\operatorname*{argmax}_v$ are uniquely solved,

and then $\mathrm{Prox}_s \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} u^* \\ v^* \end{bmatrix}$ is well-defined.

<u>Rk</u>. The saddle envelope is rooted in Moreau envelope

Moreau envelope	minimization
Saddle ...	min max

saddle " , min-max

- Moreau envelope

$$s>0, \quad f_s(x) := \min_u \overbrace{f(u) + \frac{1}{2s}\|u-x\|^2}^{h(u,x)}$$

suppose f is ρ-W.C., i.e. $\nabla^2 f \succeq -\rho I$, $\rho>0$,

then when $\frac{1}{s} > \rho$, $u^* = \arg\min_u$ $\exists!$ and is $\mathrm{Prox}_s(x) =: x_+$

Assume: f is C^2, l-smooth, ρ-WC. $\frac{1}{s} > \rho$.

Lemma (Differential characterization) f_s is C^2, and

1) $\nabla f_s(x) \overset{①}{=} \frac{1}{s}(x - x_+) \overset{②}{=} \nabla f(x_+)$ (only needs f C^1)

2) $\nabla^2 f_s(x) = \frac{1}{s}\left(I - \left(I + s\nabla^2 f(x_+)\right)^{-1}\right)$

Pf. 1) First order condition for $x_+ = u^*$:

$$\nabla f(u^*) + \frac{1}{s}(u^* - x) = 0 \quad \Rightarrow ②$$

Since $\frac{1}{s} > \rho$, $x_+(x)$ is well-defined via the eqn

$$x = x_+ + s\nabla f(x_+) \qquad (*)$$

Now $f_s(x) = h(x_+(x), x)$ $\nabla f_s(x)$ by Danskin's, or just chain rule:

$$\partial_u h(x_+, x) = 0,$$

$$\delta f_s(x) = \cancelto{0}{\partial_u h(x_+, x)}\, \delta x_+ + \partial_x h(x_+, x)\, \delta x$$

$$= \frac{1}{s}(x - x_+)\, \delta x \quad \Rightarrow ①.$$

Thm (Danskin's Thm)

$$\phi(x) = \max_y f(x,y)$$

$y \in \mathcal{Y}$

f is differentiable in x, and $\mathcal{Y}$ is compact.

If $y^*(x)$ is unique at a given x, then

$$\nabla \phi(x) = \partial_x f(x, y^*(x)).$$

2) Apply Implicit Function Thm. to (*), $x_+ = x_+(x)$

$$\delta x = \delta x_+ + s\left[\nabla^2 f(x_+)\right]\delta x_+$$

$$\Rightarrow \delta x_+ = \left[I + s\nabla^2 f(x_+)\right]^{-1}\delta x$$

By 1), $\nabla f_s(x) = \frac{1}{s}(x - x_+)$

$$\Rightarrow \nabla^2 f_s(x) = \frac{1}{s}\left(I - \frac{\partial x_+}{\partial x}\right)$$

$$= \frac{1}{s}\left(I - \left[I + s\nabla^2 f(x_+)\right]^{-1}\right) \quad \#$$

<u>Cor</u> The stationary pts of f_s are the same as those of f.

pf. WTS $\quad \nabla f_s(x) = 0 \iff \nabla f(x) = 0$

" $\Rightarrow$ " By 1) above, $x = x_+$ and $\nabla f(x_+) = 0$

$$\Rightarrow \nabla f(x) = 0$$

" $\Leftarrow$ " Recall that $x_+ = \operatorname{argmin}_u h(u, x)$

$$\partial_u h(u, x) = \nabla f(u) + \frac{1}{s}(u - x)$$

Take $u = x$ makes $\partial_u h(u, x) = 0$.

By strongly convexity of $\min_u h(u, x)$ (b/c $\frac{1}{s} > \rho$),

this means that $u = x$ is the unique u^*, i.e. $x = x_+$.

Then $\nabla f_s(x) \overset{\text{by 1)}}{=} \nabla f(x_+) = \nabla f(x) = 0. \quad \#$

<u>Cor.</u> If f is μ-S.C., $\mu > 0$, then f_s is $\left(\frac{\mu}{1+s\mu}\right)$-S.C.

pf. By lemma 2) Hessian characterization.

Rk. If f is N.C., then f_s is also N.C.

<u>lemma</u> (lower bound)

$$f_s(x) \le f(x), \ \forall x \in \mathbb{R}^n, \text{ and } "=" \text{ iff } x = x_+.$$

pf. By def. $f_s(x) = \min_u h(u,x) \le h(u=x, x) = f(x)$

Since $\arg\min_u$ $\exists !$, "=" $\Rightarrow$ $u = x$ is the minimizer

$\Rightarrow \partial_u h(u=x, x) = 0 \Rightarrow \nabla f(x) = 0$

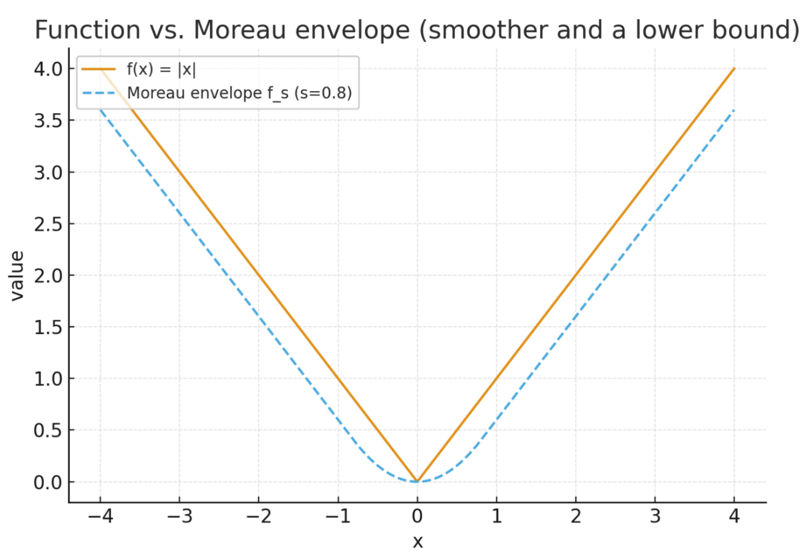
$\Rightarrow x = x_+$, similarly as above.

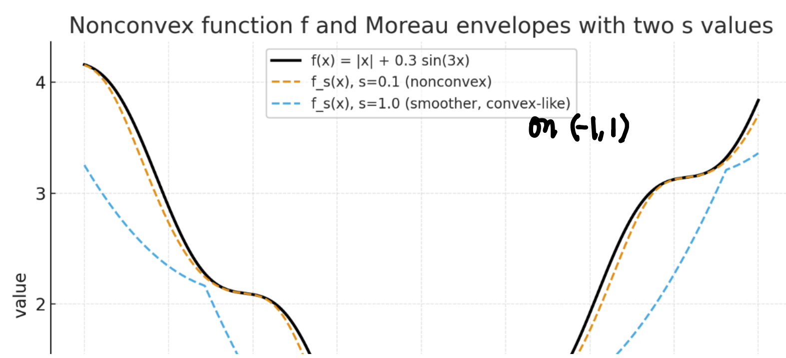
"$\Leftarrow$" If $x = x_+$, then $\nabla f(x) = 0$ by lemma 1),

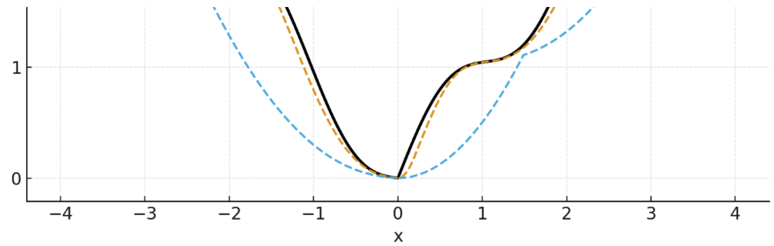
then similarly as above, $u = x$ is the minimizer u^*

$\Rightarrow f(x) = h(u^*, x) = f_s(x)$ #

Eg.







Summary : f_δ smoothes the objective, share stationary pts, provides a lower bound, but does NOT convexify f.

- Saddle envelope

$$L_\delta(x,y) = \operatorname*{argmin}_u \max_v \overbrace{L(u,v) + \frac{1}{2\delta}\|u-x\|^2 - \frac{1}{2\delta}\|v-y\|^2}^{H(u,v;x,y)}$$

L is C^2, ℓ-smooth, ρ-WC-WC, $\frac{1}{\delta} > \rho$

then $(u^*, v^*) = \operatorname*{argmin}_u \max_v$ $\exists\,!$, denote as (x_+, y_+)

We introduce notation $z = \begin{bmatrix} x \\ y \end{bmatrix}$, $z_+ = \begin{bmatrix} x_+ \\ y_+ \end{bmatrix}$.

for $f(z) = f(x,y)$, $\partial f(z) := \begin{bmatrix} \nabla_x f \\ -\nabla_y f \end{bmatrix}\Big|_z$

<u>lemma</u> 1) $L_\delta(z)$ is C^1 and

$$\partial L_\delta(z) = \frac{1}{\delta}(z - z_+) = \partial L(z_+)$$

Pf: WTS $\begin{bmatrix} \nabla_x L_\delta(x,y) \\ -\nabla_y L_\delta(x,y) \end{bmatrix} \overset{①}{=} \frac{1}{\delta}\begin{bmatrix} x - x_+ \\ y - y_+ \end{bmatrix} \overset{②}{=} \begin{bmatrix} \nabla_x L(x_+, y_+) \\ -\nabla_y L(x_+, y_+) \end{bmatrix}$

First order condition for $\min_u$, $\max_v$ give ②

To prove ①, observe that for fixed y

$$L_\delta(x,y) = \min_u \left(\max_v L(u,v) - \frac{1}{2\delta}\|v-y\|^2 \right) + \frac{1}{2\delta}\|u-x\|^2$$

$$\underbrace{\hphantom{xxxxxxxxxx}}_{g(u;y)}$$

$$= e_s\{g(\cdot;y)\}(x).$$

By the differential characterisation of Morean envelope of g,

$$\nabla_x \underbrace{e_s\{g(\cdot;y)\}(x)}_{L_s(x,y)} = \frac{1}{s}(x-x_+).$$

this is first eqn in ①.

The second eqn about ∇_y is by symmetry. #

Cor The stationary pts of L_s are the same as those of L.

pf (Ex)

lemma 2) Hessian of L_s: denote $\partial^2 f := \begin{bmatrix} \nabla_{xx}^2 f & \nabla_{xy}^2 f \\ -\nabla_{yx}^2 f & -\nabla_{yy}^2 f \end{bmatrix}$

then
$$[\partial^2 L_s(z)] = \frac{1}{s}\left(I-\left(I+s[\partial^2 L(z_+)]\right)^{-1}\right)$$

Pf. By lemma 1),

$$s\begin{bmatrix} \nabla_x L_s(x,y) \\ -\nabla_y L_s(x,y)\end{bmatrix} = \begin{bmatrix} x \\ y\end{bmatrix} - \begin{bmatrix} x_+ \\ y_+\end{bmatrix} = s\begin{bmatrix} \nabla_x L(x_+,y_+) \\ -\nabla_y L(x_+,y_+)\end{bmatrix}$$

By chain Rule,

$$s\begin{bmatrix} \nabla_{xx}^2 L_s(x,y) & \nabla_{xy}^2 L_s(x,y) \\ -\nabla_{yx}^2 L_s(x,y) & -\nabla_{yy}^2 L_s(x,y)\end{bmatrix}\begin{bmatrix}\delta x \\ \delta y\end{bmatrix} = \begin{bmatrix}\delta x \\ \delta y\end{bmatrix} - \begin{bmatrix}\delta x_+ \\ \delta y_+\end{bmatrix},$$

$$\begin{bmatrix}\delta x \\ \delta y\end{bmatrix} - \begin{bmatrix}\delta x_+ \\ \delta y_+\end{bmatrix} = s\begin{bmatrix} \nabla_{xx}^2 L(z_+) & \nabla_{xy}^2 L(z_+) \\ -\nabla_{yx}^2 L(z_+) & -\nabla_{yy}^2 L(z_+)\end{bmatrix}\begin{bmatrix}\delta x_+ \\ \delta y_+\end{bmatrix}$$

$$\Rightarrow \begin{bmatrix}\delta x_+ \\ \delta y_+\end{bmatrix} = \left(I+s[\partial^2 L(z_+)]\right)^{-1}\begin{bmatrix}\delta x \\ \delta y\end{bmatrix}$$

Then

$$s\left[\partial^2 L_s(z)\right]\delta z = \delta z - \delta z_+$$
$$= \left(I - \left(I + s\left[\partial^2 L(z_+)\right]\right)^{-1}\right)\delta z \qquad \#$$

Rk. The expression is

$$\left.\begin{bmatrix} \underline{\nabla^2_{xx} L_s} & \nabla^2_{xy} L_s \\ -\nabla^2_{yx} L_s & \underline{-\nabla^2_{yy} L_s} \end{bmatrix}\right|_z = \frac{1}{s}\left(I - \left(I + s\left.\begin{bmatrix} \nabla^2_{xx} L & \nabla^2_{xy} L \\ -\nabla^2_{yx} L & -\nabla^2_{yy} L \end{bmatrix}\right|_{z_+}\right)^{-1}\right)$$

Key idea: $\nabla^2_{xx} L_s$, $-\nabla^2_{yy} L_s$ can be PD even $\nabla^2_{xx} L$, $-\nabla^2_{yy} L$ NOT, due to the "interaction" $\nabla^2_{xy} L$

Summary: L_s smoothes the objective, share stationary pts, does NOT provide lower / upper bound,

But can convexify-concavify NC-NC L when there is large enough interaction btw x and y.