Variational Theory and Algorithms for a Class of Asymptotically Approachable Nonconvex Problems

Hanyang Li

Department of IEOR, University of California, Berkeley

Joint work with Ying Cui (UC Berkeley)

INFORMS Annual Meeting
Oct 16, 2023

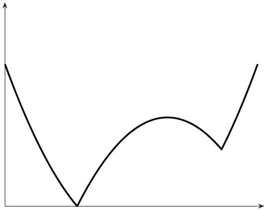
Amenable functions

Amenable: [Poliquin & Rockafellar, 1992] F has local representation

$$F = g \circ f = (\text{convex}) \circ (\text{smooth}).$$

Examples:

$$F(X) = \left\|\mathcal{A}\left(XX^\top\right) - b\right\|_1 \quad \leftarrow \quad \text{Phase retrieval, robust PCA, ...}$$



$F(x) = |ax^2 - b| + x$

Amenable functions

Amenable: [Poliquin & Rockafellar, 1992] F has local representation

$$F = g \circ f = (\text{convex}) \circ (\text{smooth}).$$

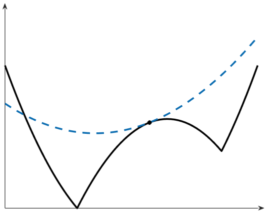
Examples:

$$F(X) = \left\|\mathcal{A}\left(XX^\top\right) - b\right\|_1 \quad \leftarrow \quad \text{Phase retrieval, robust PCA, ...}$$

Facts:

- Convex approximation

$$F_{x^k}(x) \triangleq \underbrace{g\Big(f(x^k) + \nabla f(x^k)(x - x^k)\Big)}_{(\text{convex})\circ(\text{linear})} + \frac{\|x - x^k\|^2}{2\eta}$$



$F(x) = |ax^2 - b| + x$

Amenable functions

Amenable: [Poliquin & Rockafellar, 1992] F has local representation

$$F = g \circ f = (\text{convex}) \circ (\text{smooth}).$$

Examples:

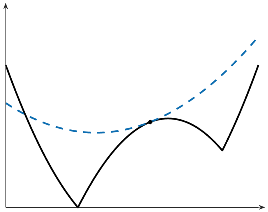
$$F(X) = \left\| \mathcal{A}(XX^\top) - b \right\|_1 \quad \leftarrow \quad \text{Phase retrieval, robust PCA, ...}$$

Facts:

- Convex approximation

$$F_{x^k}(x) \triangleq \underbrace{g\Big(f(x^k) + \nabla f(x^k)(x - x^k)\Big)}_{(\text{convex})\circ(\text{linear})} + \frac{\|x - x^k\|^2}{2\eta}$$

- Chain rule $\partial F(x) = \nabla f(x)^* \, \partial g\big(f(x)\big)$



$F(x) = |ax^2 - b| + x$

Amenable functions

Amenable: [Poliquin & Rockafellar, 1992] F has local representation

$$F = g \circ f = \text{(convex)} \circ \text{(smooth)}.$$

Examples:

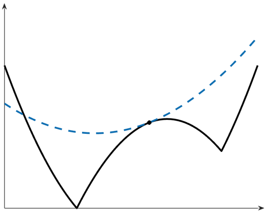
$$F(X) = \left\|\mathcal{A}\left(XX^\top\right) - b\right\|_1 \quad \leftarrow \quad \text{Phase retrieval, robust PCA, ...}$$

Facts:

- Convex approximation

$$F_{x^k}(x) \triangleq \underbrace{g\Big(f(x^k) + \nabla f(x^k)(x - x^k)\Big)}_{\text{(convex)}\circ\text{(linear)}} + \frac{\|x - x^k\|^2}{2\eta}$$

- Chain rule $\partial F(x) = \nabla f(x)^* \, \partial g\Big(f(x)\Big)$



$F(x) = |ax^2 - b| + x$

Prox-linear algorithm for $\min_x F(x) \triangleq g(f(x))$:

$$\boxed{x^{k+1} = \operatorname{argmin}_x \; F_{x^k}(x)}$$

See [Fletcher, 1982, Burke & Ferris, 1995, Lewis & Wright, 2016, Drusvyatskiy & Paquette, 2019, ...]

Amenable functions

Amenable: [Poliquin & Rockafellar, 1992] F has local representation

$$F = g \circ f = (\text{convex}) \circ (\text{smooth}).$$

Examples:

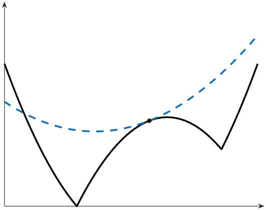
$$F(X) = \left\| \mathcal{A}\left(XX^\top\right) - b \right\|_1 \quad \leftarrow \quad \text{Phase retrieval, robust PCA, ...}$$

Facts:

- Convex approximation

$$F_{x^k}(x) \triangleq \underbrace{g\Big(f(x^k) + \nabla f(x^k)(x - x^k)\Big)}_{(\text{convex})\circ(\text{linear})} + \frac{\|x - x^k\|^2}{2\eta}$$

- Chain rule $\partial F(x) = \nabla f(x)^* \, \partial g\Big(f(x)\Big)$



$F(x) = |ax^2 - b| + x$

Prox-linear algorithm for $\min_x F(x) \triangleq g(f(x))$:

$$\boxed{x^{k+1} = \operatorname{argmin}_x \; F_{x^k}(x)}$$

Q: Can we go beyond amenable functions by relaxing the smoothness of F?

Beyond amenability

$$\min_{x \in X} F(x) \triangleq \varphi(f(x)),$$

where φ is convex and f is NOT continuously differentiable.

Motivating examples:

- Inverse optimal value problems [Ahmed & Guan, 2005]

$$f_i(x) \triangleq \begin{bmatrix} \min\limits_{y} & x^\top y \\ \text{s.t.} & A^i y \le b^i \end{bmatrix} \quad \forall\, i = 1, \cdots, m$$

  - x: unknown parameter of a linear program
  - $\{v_i\}_{1 \le i \le m}$: (noisy) observation of optimal values

  How to estimate the parameter x?

Beyond amenability

$$\min_{x \in X} F(x) \triangleq \varphi(f(x)),$$

where φ is convex and f is NOT continuously differentiable.

Motivating examples:

- Inverse optimal value problems [Ahmed & Guan, 2005]

$$f_i(x) \triangleq \begin{bmatrix} \min\limits_{y} & x^\top y \\ \text{s.t.} & A^i y \le b^i \end{bmatrix} \quad \forall\, i = 1, \cdots, m$$

  - x: unknown parameter of a linear program
  - $\{v_i\}_{1 \le i \le m}$: (noisy) observation of optimal values

  How to estimate the parameter x?

$$\min_{x \in X} \sum_{i=1}^{m} |v_i - f_i(x)| = (\text{convex}) \circ (\text{nonsmooth})$$

Beyond amenability

$$\min_{x \in X} F(x) \triangleq \varphi(f(x)),$$

where φ is convex and f is NOT continuously differentiable.

Motivating examples:

- Risk management with Value-at-risk:

$$\begin{aligned} \min_{x} \quad & \text{obj} \\ \text{s.t.} \quad & \mathrm{VaR}_{\alpha}(c(x, Z)) \geq r \end{aligned}$$

- $c(x, Z)$: the random profit of investments parameterized by decision x

Beyond amenability

$$\min_{x \in X} F(x) \triangleq \varphi(f(x)),$$

where φ is convex and f is NOT continuously differentiable.

Motivating examples:

- Risk management with Value-at-risk:

$$\begin{aligned} \min_x \ & \text{obj} \\ \text{s.t.} \ & \mathrm{VaR}_\alpha(c(x, Z)) \geq r \end{aligned}$$

- $c(x, Z)$: the random profit of investments parameterized by decision x

$$\delta_{[r,+\infty)}\Big(\mathrm{VaR}_\alpha(c(x, Z))\Big) = (\text{convex}) \circ (\text{nonsmooth})$$

$$\text{e.g. } Z = \begin{cases} 2 & \text{w.p. } 1/2 \\ 1 & \text{w.p. } 1/6 \\ -1 & \text{w.p. } 1/3 \end{cases} \quad \Rightarrow \quad \mathrm{VaR}_{1/3}(xZ + 1) = \min(2x + 1, -x + 1)$$

Asymptotic approachable problems

$$\min_{x \in X} \; F(x) \triangleq \varphi(f(x)),$$

where $\varphi : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ is convex, nondecreasing.

Structure assumption for f:

- Recall for smooth f with Lipschitz gradient:

$$f(x) - \left(f(x^k) + \nabla f(x^k)^\top (x - x^k)\right) = O(\|x - x^k\|^2)$$

Asymptotic approachable problems

$$\min_{x \in X} \; F(x) \triangleq \varphi(f(x)),$$

where $\varphi : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ is convex, nondecreasing.

Structure assumption for f:

- Recall for smooth f with Lipschitz gradient:

$$f(x) - \left(f(x^k) + \nabla f(x^k)^\top (x - x^k)\right) = O(\|x - x^k\|^2)$$

- Attempt:
  - Suppose $f = g - h$ (g, h convex).

    $\left(\text{linearization of } g \text{ at } x^k\right) - h(x) \leq f(x) \leq g(x) - \left(\text{linearization of } h \text{ at } x^k\right)$

Asymptotic approachable problems

$$\min_{x \in X} F(x) \triangleq \varphi(f(x)),$$

where $\varphi : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ is convex, nondecreasing.

Structure assumption for f:

- Recall for smooth f with Lipschitz gradient:

$$f(x) - \left(f(x^k) + \nabla f(x^k)^\top (x - x^k)\right) = O(\|x - x^k\|^2)$$

- Attempt:
  - Suppose $f = g - h$ (g, h convex).

$$\left(\text{linearization of } g \text{ at } x^k\right) - h(x) \le f(x) \le g(x) - \left(\text{linearization of } h \text{ at } x^k\right)$$

Limitation: only includes locally Lipschitz functions!

Asymptotic approachable problems

$$\min_{x \in X} \; F(x) \triangleq \varphi(f(x)),$$

where $\varphi : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ is convex, nondecreasing.

Structure assumption for f:

- Recall for smooth f with Lipschitz gradient:

$$f(x) - \left(f(x^k) + \nabla f(x^k)^\top (x - x^k)\right) = O(\|x - x^k\|^2)$$

- Attempt:
  - ~~Suppose $f = g - h$ (g, h convex).~~

    $$\left(\text{linearization of } g \text{ at } x^k\right) - h(x) \le f(x) \le g(x) - \left(\text{linearization of } h \text{ at } x^k\right)$$

    Limitation: only includes locally Lipschitz functions!

  - Suppose f is approachable difference-of-convex (ADC):

    $$f = \lim_k \left(f^k \triangleq g^k - h^k\right) \qquad g^k, h^k \text{ convex.}$$

Asymptotic approachable problems

$$\min_{x \in X} F(x) \triangleq \varphi(f(x)) = (\text{convex nondecreasing}) \circ (\text{ADC}).$$

Ubiquity of ADC:

Theorem ([Royset, 2020])

$\forall$ lower semicontinuous function f, $\exists\{f^k\}$ such that $f = (\text{epi-})\lim_k f^k$, where

$$f^k(x) = \underbrace{\max_{1 \le i \le p_k} \left[\langle a^{k,i}, x\rangle + \alpha_i\right]}_{\text{convex}} - \underbrace{\max_{1 \le i \le q_k} \left[\langle b^{k,i}, x\rangle + \beta_i\right]}_{\text{convex}} \qquad \forall\, k.$$

Asymptotic approachable problems

$$\min_{x\in X} F(x) \triangleq \varphi(f(x)) = (\text{convex nondecreasing}) \circ (\text{ADC}).$$

Ubiquity of ADC:

- Observation: Moreau envelope

$$e_{\lambda_k} f(x) \quad = \inf_u \left\{ f(u) + \frac{\|u-x\|^2}{2\lambda_k} \right\}$$

Asymptotic approachable problems

$$\min_{x \in X} F(x) \triangleq \varphi(f(x)) = (\text{convex nondecreasing}) \circ (\text{ADC}).$$

Ubiquity of ADC:

- Observation: Moreau envelope

$$\begin{aligned} e_{\lambda_k} f(x) &= \inf_u \left\{ f(u) + \frac{\|u - x\|^2}{2\lambda_k} \right\} \\ &= \frac{\|x\|^2}{2\lambda_k} - \underbrace{\sup_u \left\{ -f(u) - \frac{\|u\|^2 - 2u^\top x}{2\lambda_k} \right\}}_{\text{convex in } x} \end{aligned}$$

Asymptotic approachable problems

$$\min_{x \in X} F(x) \triangleq \varphi(f(x)) = (\text{convex nondecreasing}) \circ (\text{ADC}).$$

Ubiquity of ADC:

- Observation: Moreau envelope

$$\begin{aligned} e_{\lambda_k} f(x) &= \inf_u \left\{ f(u) + \frac{\|u - x\|^2}{2\lambda_k} \right\} \\ &= \frac{\|x\|^2}{2\lambda_k} - \underbrace{\sup_u \left\{ -f(u) - \frac{\|u\|^2 - 2u^\top x}{2\lambda_k} \right\}}_{\text{convex in } x} \end{aligned}$$

$\Longrightarrow$ Any lower semicontinuous, prox-bounded function f is ADC.

Asymptotic approachable problems

$$\min_{x \in X} F(x) \triangleq \varphi(f(x)) = (\text{convex nondecreasing}) \circ (\text{ADC}).$$

Ubiquity of ADC:

- Observation: Moreau envelope

$$\begin{aligned} e_{\lambda_k} f(x) &= \inf_u \left\{ f(u) + \frac{\|u - x\|^2}{2\lambda_k} \right\} \\ &= \frac{\|x\|^2}{2\lambda_k} - \underbrace{\sup_u \left\{ -f(u) - \frac{\|u\|^2 - 2u^\top x}{2\lambda_k} \right\}}_{\text{convex in } x} \end{aligned}$$

$\Longrightarrow$ Any lower semicontinuous, prox-bounded function f is ADC.

- More examples: the optimal value functions, value-at-risk are nonsmooth but ADC!

Convex approximation and algorithms

$$\min_{x \in X} F(x) \triangleq \varphi(f(x)) = (\text{convex nondecreasing}) \circ (\text{ADC}).$$

Need to solve

$$\min_{x \in X} F^k(x) \triangleq \varphi\Big(\underbrace{g^k(x) - h^k(x)}_{= f^k(x)} \Big) \qquad \forall\, k.$$

Convex approximation and algorithms

$$\min_{x \in X} \; F(x) \triangleq \varphi(f(x)) = (\text{convex nondecreasing}) \circ (\text{ADC}).$$

Need to solve

$$\min_{x \in X} \; F^k(x) \triangleq \varphi\Big(\underbrace{g^k(x) - h^k(x)}_{= f^k(x)} \Big) \qquad \forall\, k.$$

Convex approximation:

$$\underbrace{\varphi\Big(g^k(x) - \text{linearization of } h^k}_{(\text{nondecreasing}) \,\circ\, (\text{convex})} \Big) \geq \varphi\big(f^k(x)\big)$$

Convex approximation and algorithms

$$\min_{x \in X} \; F(x) \triangleq \varphi(f(x)) = (\text{convex nondecreasing}) \circ (\text{ADC}).$$

Need to solve

$$\min_{x \in X} \; F^k(x) \triangleq \varphi\Big(\underbrace{g^k(x) - h^k(x)}_{=f^k(x)} \Big) \qquad \forall\, k.$$

Convex approximation:

$$\underbrace{\varphi\Big(g^k(x) - \text{linearization of } h^k}_{(\text{nondecreasing}) \,\circ\, (\text{convex})} \Big) \geq \varphi\big(f^k(x)\big)$$

Algorithm framework:

Solve $\min_x F^k(x)$ iteratively by its convex approximation

Convex approximation and algorithms

$$\min_{x \in X} \; F(x) \triangleq \varphi(f(x)) = (\text{convex nondecreasing}) \circ (\text{ADC}).$$

Need to solve

$$\min_{x \in X} \; F^k(x) \triangleq \varphi\Big(\underbrace{g^k(x) - h^k(x)}_{= f^k(x)} \Big) \qquad \forall\, k.$$

Convex approximation:

$$\underbrace{\varphi\Big(g^k(x) - \text{linearization of } h^k \Big)}_{(\text{nondecreasing}) \circ (\text{convex})} \geq \varphi\big(f^k(x)\big)$$

Algorithm framework:

For $k = 1, 2, \cdots$

Solve $\min_x F^k(x)$ iteratively by its convex approximation

(within proper tolerance $\varepsilon^k \downarrow 0$)

Asymptotic stationary

$$\min_{x \in X} F(x) \triangleq \varphi(f(x)) = (\text{convex nondecreasing}) \circ (\text{ADC}) \qquad \text{(P)}$$

Convergence: Any accumulation point $\bar{x}$ generated by the algorithm satisfies

$$0 \in \partial\varphi\big(f(\bar{x})\big) \cdot \partial_A f(\bar{x}),$$ [1]

where $\partial_A f(x)$ is a "subdifferential" of f dependent on $\{f^k = g^k - h^k\}$:

$$\partial_A f(x) \triangleq \bigcup_{x^k \to x} \left\{\text{accumulation points of } \left[\partial g^k(x^k) - \partial h^k(x^k)\right]\right\}.$$

[1] This generalizes the stationary point in [Royset, 2022]

Asymptotic stationary

$$\min_{x \in X} \; F(x) \triangleq \varphi(f(x)) = (\text{convex nondecreasing}) \circ (\text{ADC}) \qquad \text{(P)}$$

Convergence: Any accumulation point $\bar{x}$ generated by the algorithm satisfies

$$0 \in \partial\varphi\big(f(\bar{x})\big) \cdot \partial_A f(\bar{x}) \,, \,^{1}$$

where $\partial_A f(x)$ is a "subdifferential" of f dependent on $\{f^k = g^k - h^k\}$:

$$\partial_A f(x) \triangleq \bigcup_{x^k \to x} \left\{\text{accumulation points of } \left[\partial g^k(x^k) - \partial h^k(x^k)\right]\right\}.$$

Q: Is it a necessary condition for a local minimizer of (P)?

[1] This generalizes the stationary point in [Royset, 2022]

Good or bad approximation? —Epi-convergence

Epi-convergence:

$$f^k \xrightarrow{e} f \quad \iff \quad \mathrm{Lim}_k \left(\mathrm{epi}\, f^k\right) = \mathrm{epi}\, f$$

[2] See monographs [Rockafellar & Wets, 2009], [Royset & Wets, 2022]

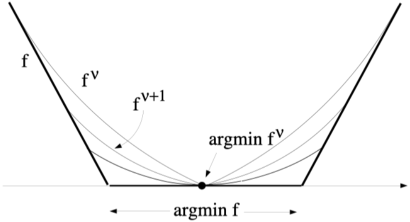
Good or bad approximation? —Epi-convergence

Epi-convergence:

$$f^k \xrightarrow{e} f \quad \Longleftrightarrow \quad \mathrm{Lim}_k \left(\mathrm{epi}\, f^k\right) = \mathrm{epi}\, f$$

Fact[2]**:** (global minimizers)

$$f^k \xrightarrow{e} f \quad \Longrightarrow \quad \left\{\text{accumulation points of } \left(\varepsilon^k\text{-argmin}\, f^k\right)\right\} \subset \mathrm{argmin}\, f, \quad \forall \varepsilon^k \downarrow 0.$$



[2] See monographs [Rockafellar & Wets, 2009], [Royset & Wets, 2022]

Necessary optimality condition under epi-convergence

$$\min_{x \in X} \; F(x) \triangleq \varphi(f(x)) = (\text{convex nondecreasing}) \circ (\text{ADC}) \qquad \text{(P)}$$

Convergence: Any accumulation point $\bar{x}$ generated by the algorithm satisfies

$$0 \in \partial\varphi\big(f(\bar{x})\big) \cdot \partial_A f(\bar{x}) \,,$$

where $\partial_A f(x)$ is a "subdifferential" of f dependent on $\{f^k = g^k - h^k\}$.

Q: Is it a necessary condition for a local minimizer of (P)?

A: Under $f^k \xrightarrow{e} f$!

[3] NOT computationally friendly

Necessary optimality condition under epi-convergence

$$\min_{x \in X} \; F(x) \triangleq \varphi(f(x)) = (\text{convex nondecreasing}) \circ (\text{ADC}) \qquad \text{(P)}$$

Convergence: Any accumulation point $\bar{x}$ generated by the algorithm satisfies

$$0 \in \partial\varphi\big(f(\bar{x})\big) \cdot \partial_A f(\bar{x}),$$

where $\partial_A f(x)$ is a "subdifferential" of f dependent on $\{f^k = g^k - h^k\}$.

Q: Is it a necessary condition for a local minimizer of (P)?

A: Under $f^k \xrightarrow{e} f$!

Idea:

Any local minimizer of (P) $\Longrightarrow$ A stationary point[3] for (P)

$\Updownarrow$

$$0 \in \partial\varphi\big(f(\bar{x})\big) \cdot \partial_A f(\bar{x})$$

[3] NOT computationally friendly

Extensions

Our analysis can be applied to

$$\min_x \sum (\text{univariate convex}) \circ (\text{ADC}).$$

Lemma (a monotonic decomposition)

For a univariate convex function φ, $\exists$ convex $\varphi^{\uparrow}$ and $\varphi^{\downarrow}$ such that

$$\varphi = \varphi^{\uparrow}(\text{nondecreasing}) + \varphi^{\downarrow}(\text{nonincreasing}).$$

Conclusion

- Beyond amenability: A class of asymptotically approachable problems

$$\min_x \sum (\text{univariate convex}) \circ (\text{ADC})$$

- Epi-convergence:

$$\Longrightarrow \quad \text{consistency of the global min} + \underbrace{\text{stationary points (subgradient-based)}}_{\text{necessary for the local min}}$$

Thank you!

This talk is based on the work:

- Hanyang Li, Ying Cui. *Variational Theory and Algorithms for a Class of Asymptotically Approachable Nonconvex Problems*. arXiv:2307.00780 (2023)

Appendix 1: Subgradient relationship under epi-convergence

Proposition

If f is ADC and $f^k (= g^k - h^k) \xrightarrow{e} f$, then

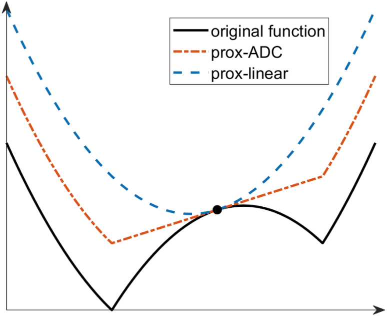
(a) $\partial f(x) \subset \partial_A f(x)$;

(b) $\partial f(x) \subset \partial_A f(x) \subset \operatorname{conv} \partial f(x)$ if f is locally Lipschitz continuous and $f^k = e_{\lambda_k} f$ for any $\lambda_k \downarrow 0$.

Appendix 2: Comparison with prox-linear method

$$\min_x \ \varphi(f(x))$$

φ is univariate convex, f is smooth with ℓ-Lipschitz gradient



$$f = \underbrace{\left(f + \frac{\ell}{2}\| \bullet \|^2\right)}_{\text{convex}} - \frac{\ell}{2}\| \bullet \|^2$$

Prox-ADC:

$$\varphi^{\uparrow}\left(f(x) + \frac{\ell}{2}\|x - x^k\|^2\right) + \varphi^{\downarrow}\left(f(x^k) + \nabla f(x^k)^\top (x - x^k) - \frac{\ell}{2}\|x - x^k\|^2\right)$$

Prox-linear:

$$\varphi\big(f(x^k) + \nabla f(x^k)^\top (x - x^k)\big)$$