

Introductory Course on Non-smooth Optimisation

Lecture 09 - Non-convex optimisation

Jingwei Liang

Department of Applied Mathematics and Theoretical Physics

- 1 Examples
- 2 Non-convex optimisation
- 3 Convex relaxation
- 4 Łojasiewicz inequality
- 5 Kurdyka-Łojasiewicz inequality

Forward observation

$$b = A\hat{x},$$

- $\hat{x} \in \mathbb{R}^n$ is sparse.
- $A : \mathbb{R}^n \rightarrow \mathbb{R}^m$ with $m \ll n$.

Compressed sensing

$$\min_{x \in \mathbb{R}^n} \|x\|_0 \quad \text{s.t.} \quad Ax = b.$$

NB: NP-hard problem.

Two-phase segmentation Given an image I , which consists of foreground and background, segment the foreground. Ideally,

$$I = f_C + b_{\Omega \setminus C}.$$

Mumford-Shah model

$$E(u, C) = \int_{\Omega} (u - I)^2 dx + \lambda \left(\int_{\Omega \setminus C} \|\nabla u\|^2 dx + \alpha |C| \right),$$

where $|C| = \text{peri}(C)$.

Forward mixture model

$$w = \hat{x} + \hat{y} + \epsilon,$$

where $\hat{x} \in \mathbb{R}^{m \times n}$ is κ -sparse, $\hat{y} \in \mathbb{R}^{m \times n}$ is σ -low-rank and ϵ is noise.

Non-convex PCP

$$\begin{aligned} \min_{x, y \in \mathbb{R}^{m \times n}} \quad & \frac{1}{2} \|x + y - w\|^2 \\ \text{s.t.} \quad & \|x\|_0 \leq \kappa \quad \text{and} \quad \text{rank}(y) \leq \sigma. \end{aligned}$$

Each layer of NNs is convex

- Linear operation, e.g. convolution.
- Non-linear activation function, e.g. rectifier $\max\{x, 0\}$.

The composition of convex functions is not necessarily convex...

- Neural networks are universal function approximators.
- Hence need to approximate non-convex functions.
- Cannot approximate non-convex functions with convex functions.

1 Examples

2 Non-convex optimisation

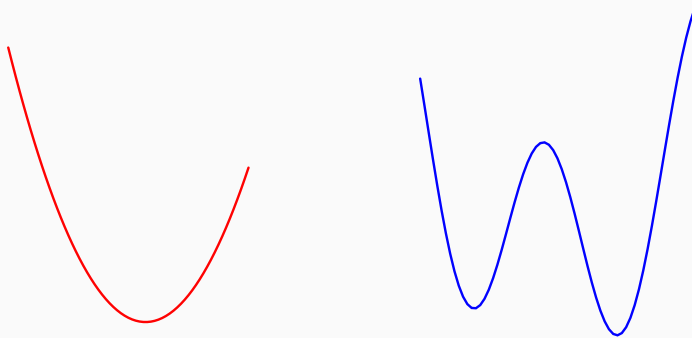
3 Convex relaxation

4 Łojasiewicz inequality

5 Kurdyka-Łojasiewicz inequality

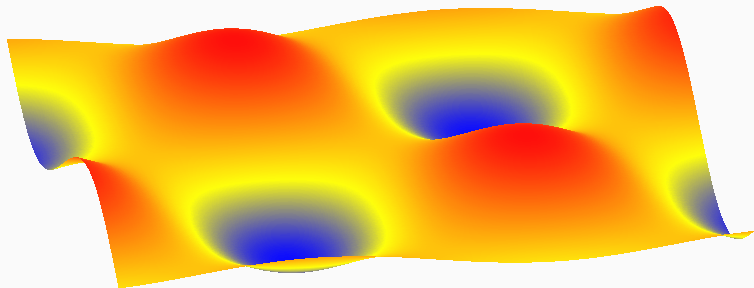
Non-convex problem

Any problem that is not convex/concave is non-convex...



Challenges

- Potentially many local minima.
- Saddle points.
- Very flat regions.
- Widely varying curvature.
- NP-hard.



1 Examples

2 Non-convex optimisation

3 Convex relaxation

4 Łojasiewicz inequality

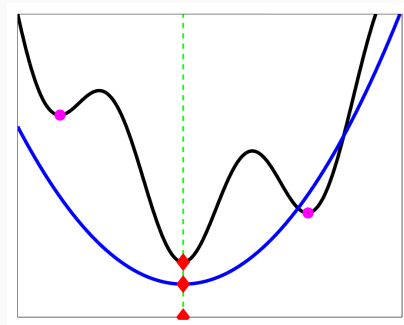
5 Kurdyka-Łojasiewicz inequality

Non-convex optimisation problem

$$\min_x E(x).$$

Convex optimisation problem

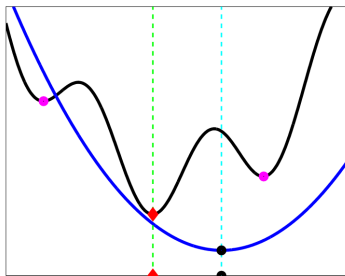
$$\min_x F(x).$$



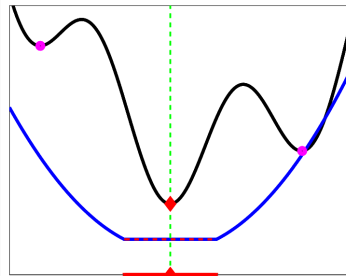
What if

$$\text{Argmin}(F) \subseteq \text{Argmin}(E),$$

- Subtle and case-dependent.
- Somehow, finding F is almost equivalent to solving E .



Loose relaxation



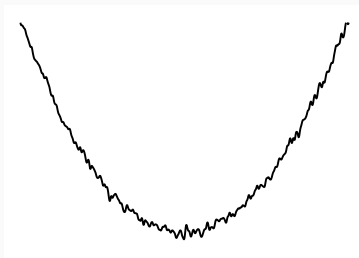
Ideal relaxation

- In practice, it is easier to obtain

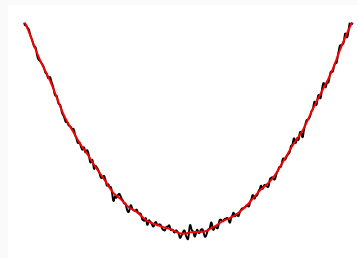
$$\text{Argmin}(E) \subseteq \text{Argmin}(F).$$

- Loose relaxation will **work** if two global minima are close enough.
- Ideal relaxation will **fail** if $\text{Argmin}(F)$ is too large.

For certain problems, non-convexity can be treated as noise...



Original function



Convolution

- Symmetric boundary condition for the convolution.
- Almost convex problem after convolution.

- 1 Examples
- 2 Non-convex optimisation
- 3 Convex relaxation
- 4 Łojasiewicz inequality**
- 5 Kurdyka-Łojasiewicz inequality

Let $F \in C_L^1$.

- Gradient descent

$$x_{k+1} = x_k - \gamma \nabla F(x_k).$$

- Descent property

$$F(x_k) - F(x_{k+1}) \geq \gamma(1 - \frac{\gamma L}{2}) \|\nabla F(x_k)\|^2.$$

- Let $\gamma \in]0, 2/L[$,

$$\gamma(1 - \frac{\gamma L}{2}) \sum_{i=0}^k \|\nabla F(x_i)\|^2 \leq F(x_0) - F(x_{k+1}) \leq F(x_0) - F(x^*).$$

- $F(x^*) > -\infty$, rhs is a positive constant.

- for lhs, let $k \rightarrow +\infty$,

$$\lim_{k \rightarrow +\infty} \|\nabla F(x_k)\|^2 = 0.$$

NB: for smooth case, a critical point is guarantee. For non-smooth problem...

Semi-algebraic set

A semi-algebraic subset of \mathbb{R}^n is a finite union of sets of the form

$$\{x \in \mathbb{R}^n : f_i(x) = 0, g_j(x) \leq 0, i \in I, j \in J\}$$

where I, J are finite and $f_i, g_j : \mathbb{R}^n \rightarrow \mathbb{R}$ are real polynomial functions.

- Stability under finite \cap, \cup and complementation.

Semi-algebraic set

A function or a mapping is semi-algebraic if its graph is a semi-algebraic set.

- Same definition for real-extended function or multivalued mappings.

Tarski-Seidenberg

The image of a semi-algebraic set by a linear projection is semi-algebraic.

- The closure of a semi-algebraic set A is semi-algebraic.

Example

- The graph of the derivative of a semi-algebraic function is semi-algebraic.
- Let A be a semi-algebraic subset of \mathbb{R}^n and $f : \mathbb{R}^n \rightarrow \mathbb{R}^p$ semi-algebraic. Then $f(A)$ is semi-algebraic.
- $g(x) = \max\{F(x, y) : y \in S\}$ is semi-algebraic if F and S are semi-algebraic.
- Other examples

$$\min_x \frac{1}{2} \|Ax - b\|^2 + \mu \|x\|_p : p \text{ is rational},$$

$$\min_x \frac{1}{2} \|AX - B\|^2 + \mu \text{rank}(X).$$

Convex subdifferential $R \in \Gamma_0(\mathbb{R}^n)$

$$\partial R(x) = \{g : R(x') \geq R(x) + \langle g, x' - x \rangle, \forall x' \in \mathbb{R}^n\}.$$

Fréchet subdifferential

Given $x \in \text{dom}(R)$, the Fréchet subdifferential $\hat{\partial}R(x)$ of R at x is the set of vectors v such that

$$\liminf_{x' \rightarrow x, x' \neq x} \frac{1}{\|x - x'\|} (R(x') - R(x) - \langle v, x' - x \rangle) \geq 0.$$

- If $x \notin \text{dom}(R)$, then $\hat{\partial}R(x) = \emptyset$.

Limiting subdifferential

The limiting-subdifferential (or simply subdifferential) of R at x , written as $\partial R(x)$, reads

$$\partial R(x) \stackrel{\text{def}}{=} \{v \in \mathbb{R}^n : \exists x_k \rightarrow x, R(x_k) \rightarrow R(x), v_k \in \hat{\partial}R(x_k) \rightarrow v\}.$$

- $\hat{\partial}R$ is convex and ∂R is closed.

Minimal norm subgradient

$$\|\partial R(x)\|_- = \min\{\|v\| : v \in \partial R(x)\}.$$

Critical points

- Fermat's rule: if x is a minimiser of R , then $0 \in \partial R(x)$.
- Conversely when $0 \in \partial R(x)$, the point x is called a critical point.
- When R is convex, any minimiser is a global minimiser.
- When R is non-convex
 - Local minima.
 - Local maxima.
 - Saddle point.

Sharpness

Function $R : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ is called sharp on the slice

$$[a < R < b] \stackrel{\text{def}}{=} \{x \in \mathbb{R}^n : a < f(x) < b\}.$$

If there exists $\alpha > 0$ such that

$$\|\partial R(x)\|_- \geq \alpha, \forall x \in [a < R < b].$$

- Norms, e.g. $R(x) = \|x\|$.



Łojasiewicz inequality

Let $R : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ be proper lower semi-continuous, and moreover continuous along its domain. Then R is said to have Łojasiewicz property if: for any critical point \bar{x} , there exist $C, \epsilon > 0$ and $\theta \in [0, 1[$ such that

$$|R(x) - R(\bar{x})|^\theta \leq C\|v\|, \quad \forall x \in \mathbb{B}_{\bar{x}}(\epsilon), v \in \partial R(x).$$

- By convention, let $0^0 = 0$.

Property

Suppose that R has Łojasiewicz property.

- If S is a connected subset of the set of critical points of R , that is $0 \in \partial R(x)$ for all $x \in S$, then R is constant on S .
- If in addition S is a compact set, then there exist $C, \epsilon > 0$ and $\theta \in [0, 1[$ such that

$$\forall x \in \mathbb{R}^n, \text{dist}(x, S) \leq \epsilon, \forall v \in \partial R(x) : |R(x) - R(\bar{x})|^\theta \leq C\|v\|.$$

Proximal point algorithm

Let $R : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ be proper and lower semi-continuous. From arbitrary $x_0 \in \mathbb{R}^n$,

$$x_{k+1} \in \operatorname{argmin}_x \gamma R(x) + \frac{1}{2} \|x - x_k\|^2.$$

Assumption

- R is proper, that is

$$\inf_{x \in \mathbb{R}^n} R(x) > -\infty.$$

This implies

$$\operatorname{argmin}_x \gamma R(x) + \frac{1}{2} \|x - x_k\|^2$$

is non-empty and compact.

- The restriction of R to its domain is a continuous function.
- R has the Łojasiewicz property.

Property

Let $\{x_k\}_{k \in \mathbb{N}}$ be the sequence generated by non-convex PPA and $\omega(x_k)$ the set of its limiting points. Then

- Sequence $\{R(x_k)\}_{k \in \mathbb{N}}$ is decreasing.
- $\sum_k \|x_k - x_{k+1}\|^2 < +\infty$.
- If R satisfies assumption 2, then $\omega(x_k) \subset \text{crit}(R)$.

If moreover, $\{x_k\}_{k \in \mathbb{N}}$ is bounded

- $\omega(x_k)$ is a non-empty compact set, and

$$\text{dist}(x_k, \omega(x_k)) \rightarrow 0.$$

- If R satisfies assumption 2, then R is finite and constant on $\omega(x_k)$.

NB: Boundedness can be guaranteed if R is coercive.

Convergence of PPA

Suppose the sequence $\{x_k\}_{k \in \mathbb{N}}$ generated by non-convex PPA is bounded, then

$$\sum_k \|x_k - x_{k+1}\| < +\infty,$$

and the whole sequence converges to some critical point $\bar{x} \in \text{crit}(R)$.

- From definition of x_{k+1} : $R(x_{k+1}) + \frac{1}{2\gamma} \|x_k - x_{k+1}\|^2 \leq R(x_k)$.

- Consider $g(s) = s^{1-\theta}$, $s > 0$: $\nabla g(s) = (1-\theta)s^{-\theta}$

$$\begin{aligned} g(R(x_k)) - g(R(x_{k+1})) &\geq (1-\theta)(R(x_{k+1}))^{-\theta} (R(x_k) - R(x_{k+1})) \\ &\geq (1-\theta)(R(x_k))^{-\theta} \frac{1}{2\gamma} \|x_k - x_{k+1}\|^2. \end{aligned}$$

- WLOG, assume $R(\bar{x}) = 0$ for $\bar{x} \in \omega(x_k)$. Let $v_k \in \partial R(x_k)$, then for all k large enough

$$0 < R(x_k)^\theta \leq C \|v_k\| = \frac{C}{\gamma} \|x_k - x_{k-1}\|.$$

- There exists $M > 0$

$$\frac{\|x_k - x_{k+1}\|^2}{\|x_k - x_{k-1}\|} \leq M (R(x_k)^{1-\theta} - R(x_{k+1})^{1-\theta}).$$

Convergence of PPA

Suppose the sequence $\{x_k\}_{k \in \mathbb{N}}$ generated by non-convex PPA is bounded, then

$$\sum_k \|x_k - x_{k+1}\| < +\infty,$$

and the whole sequence converges to some critical point $\bar{x} \in \text{crit}(R)$.

- Take $r \in]0, 1[$, if $\|x_k - x_{k+1}\| \geq r\|x_k - x_{k-1}\|$, then

$$\|x_k - x_{k+1}\| \leq \frac{M}{r} (R(x_k)^{1-\theta} - R(x_{k+1})^{1-\theta}).$$

- For all k large enough

$$\|x_k - x_{k+1}\| \leq r\|x_k - x_{k-1}\| + \frac{M}{r} (R(x_k)^{1-\theta} - R(x_{k+1})^{1-\theta}).$$

- There exists some $K > 0$, such that for $k \geq K$

$$\sum_{i=K}^k \|x_i - x_{i+1}\| \leq \frac{r}{1-r} \|x_K - x_{K-1}\| + \frac{M}{r(1-r)} (R(x_K)^{1-\theta} - R(x_{k+1})^{1-\theta}).$$

- $R(x)$ is bounded from below. Take $k \rightarrow +\infty \dots$

Convergence rate

Suppose the convergence of the non-convex PPA is true. Denote θ the Łojasiewicz exponent of x_∞ . The following statements hold

- If $\theta = 0$, then $\{x_k\}_{k \in \mathbb{N}}$ converges in finite number of steps.
- If $\theta \in]0, 1/2]$, then there exists $\eta \in]0, 1[$ such that

$$\|x_k - x_\infty\| = O(\eta^k).$$

- If $\theta \in]1/2, 1[$, then

$$\|x_k - x_\infty\| = O(k^{-\frac{1-\theta}{2\theta-1}}).$$

- 1 Examples
- 2 Non-convex optimisation
- 3 Convex relaxation
- 4 Łojasiewicz inequality
- 5 Kurdyka-Łojasiewicz inequality**

Let $R : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ be proper l.s.c. For a, b such that $-\infty < a < b < +\infty$,

$$[a < R < b] \stackrel{\text{def}}{=} \{x \in \mathbb{R}^n : a < R(x) < b\}.$$

Kurdyka-Łojasiewicz inequality

R is said to have the KL property at $\bar{x} \in \text{dom}(R)$ if there exists $\eta \in]0, +\infty]$, a neighbourhood U of \bar{x} and a continuous **concave** function $\varphi : [0, \eta[\rightarrow \mathbb{R}_+$ such that

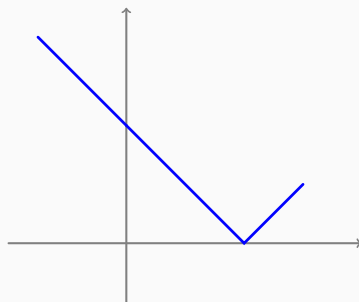
- $\varphi(0) = 0$.
- φ is C^1 on $]0, \eta[$.
- for all $s \in]0, \eta[$, $\varphi'(s) > 0$.
- for all $x \in U \cap [R(\bar{x}) < R < R(\bar{x}) + \eta]$, the KL inequality holds

$$\varphi'(R(x) - R(\bar{x})) \text{dist}(0, \partial R(x)) \geq 1.$$

- Proper l.s.c. functions are KL at non-critical points.
- Proper l.s.c. functions which satisfy KL at each point of $\text{dom}(\partial R)$ are called KL functions.
- Typical KL functions are the class of semi-algebraic functions.



$$\|\nabla F(x)\| \geq 0$$



$$\|\nabla(\varphi \circ F)(x)\| \geq 1$$

- When $R(\bar{x}) = 0$, then the condition becomes

$$\|\partial(\varphi \circ F)(x)\|_- \geq 1.$$

- φ is called a desingularising function for R , i.e. sharp up to reparameterization via φ .

Let Φ be proper and lower semi-continuous. Suppose a sequence $\{x_k\}_{k \in \mathbb{N}}$ is generated such that the following conditions are satisfied.

Conditions

Let $c, d > 0$ be some constants

A.1 Sufficient decrease conditions For each $k \in \mathbb{N}$,

$$\Phi(x_{k+1}) + c\|x_{k+1} - x_k\|^2 \leq \Phi(x_k).$$

A.2 Relative error condition For each $k \in \mathbb{N}$, there exists $g_{k+1} \in \partial\Phi(x_{k+1})$ such that

$$\|g_{k+1}\| \leq d\|x_{k+1} - x_k\|.$$

A.3 Continuity condition There exists a subsequence $\{x_{k_j}\}_{j \in \mathbb{N}}$ and \bar{x} such that

$$x_{k_j} \rightarrow \bar{x}, \quad \Phi(x_{k_j}) \rightarrow \Phi(\bar{x}).$$

Convergence

Let $\Phi : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ be proper and l.s.c. and KL at some $\bar{x} \in \mathbb{R}^n$. Let U, η and φ be in the KL property. Let $\delta, \rho > 0$ be such that $\mathbb{B}_{\bar{x}}(\delta) \subset U$ with $\rho \in]0, \delta[$. Consider a sequence $\{x_k\}_{k \in \mathbb{N}}$ which satisfies (A.1)-(A.2). Suppose moreover

$$\Phi(\bar{x}) < \Phi(x_0) < \Phi(\bar{x}) + \eta,$$

$$\|x_0 - \bar{x}\| + 2\sqrt{\frac{\Phi(x_0) - \Phi(\bar{x})}{c}} + \frac{d}{c}\varphi(\Phi(x_0) - \Phi(\bar{x})) < \rho,$$

and

$$\forall k \in \mathbb{N}, x_k \in \mathbb{B}_{\bar{x}}(\rho) \Rightarrow x_{k+1} \in \mathbb{B}_{\bar{x}}(\delta) \text{ with } \Phi(x_{k+1}) \geq \Phi(\bar{x}).$$

Then the sequence $\{x_k\}_{k \in \mathbb{N}}$ satisfies

$$\forall k \in \mathbb{N}, x_k \in \mathbb{B}_{\bar{x}}(\delta),$$

$$\sum_k \|x_k - x_{k+1}\| < +\infty,$$

$$\Phi(x_k) \rightarrow \Phi(\bar{x}).$$

and converges to a point $x^* \in \mathbb{B}_{\bar{x}}(\delta)$ such that $\Phi(x^*) \leq \Phi(\bar{x})$.

If moreover, (A.3) is true, then x^* is a critical point and $\Phi(x^*) = \Phi(\bar{x})$.

- Condition (A.1) implies that $\{\Phi(x_k)\}_{k \in \mathbb{N}}$ is non-increasing, and for all $k \in \mathbb{N}$

$$\|x_{k+1} - x_k\| \leq \sqrt{\frac{\Phi(x_k) - \Phi(x_{k+1})}{c}}.$$

- Condition (A.2) and KL inequality

$$\varphi'(\Phi(x_k) - \Phi(\bar{x})) \geq \frac{1}{\|g_k\|} \geq \frac{1}{d\|x_k - x_{k-1}\|}.$$

- Since φ is concave,

$$\begin{aligned} \varphi(\Phi(x_k) - \Phi(\bar{x})) - \varphi(\Phi(x_k) - \Phi(\bar{x})) &\geq \varphi'(\Phi(x_k) - \Phi(\bar{x}))(\Phi(x_k) - \Phi(x_{k+1})) \\ &\geq \varphi'(\Phi(x_k) - \Phi(\bar{x}))c\|x_k - x_{k+1}\|^2. \end{aligned}$$

- Combining the above two yields

$$\frac{\|x_k - x_{k+1}\|^2}{\|x_k - x_{k-1}\|} \leq \frac{d}{c}(\varphi(\Phi(x_k) - \Phi(\bar{x})) - \varphi(\Phi(x_k) - \Phi(\bar{x}))).$$

- Apply the inequality $2\sqrt{xy} \leq x + y$,

$$2\|x_k - x_{k+1}\| \leq \|x_k - x_{k-1}\| + \frac{d}{c}(\varphi(\Phi(x_k) - \Phi(\bar{x})) - \varphi(\Phi(x_{k+1}) - \Phi(\bar{x}))).$$

- Continue with (A.1),

$$\|x_1 - x_0\| \leq \sqrt{\frac{\Phi(x_0) - \Phi(x_1)}{c}} \leq \sqrt{\frac{\Phi(x_0) - \Phi(\bar{x})}{c}}.$$

- Then

$$\|x_1 - \bar{x}\| \leq \|x_1 - x_0\| + \|x_0 - \bar{x}\| \leq \|x_1 - x_0\| + \sqrt{\frac{\Phi(x_0) - \Phi(\bar{x})}{c}} \leq \rho.$$

- By induction, we can show that for all $k \in \mathbb{N}$

$$x_k \in \mathbb{B}_{\bar{x}}(\rho) \quad \text{and}$$

$$\sum_{i=1}^k \|x_{i+1} - x_i\| + \|x_{k+1} - x_k\| \leq \|x_1 - x_0\| + \frac{d}{c} (\varphi(\Phi(x_1) - \Phi(\bar{x})) - \varphi(\Phi(x_{k+1}) - \Phi(\bar{x}))).$$

- The above directly implies

$$\sum_k \|x_k - x_{k+1}\| \leq \|x_1 - x_0\| + \frac{d}{c} \varphi(\Phi(x_1) - \Phi(\bar{x})) < +\infty.$$

- Hence, there exists $x^* \in \omega(x_k)$

$$x_k \rightarrow x^*, \quad g_k \rightarrow 0, \quad \Phi(x_k) \rightarrow v \geq \Phi(\bar{x}).$$

- KL inequality

$$\varphi'(v - \Phi(\bar{x})) \|g_k\| \geq 1$$

indicates $v = \Phi(\bar{x})$. Lower semi-continuous yields $\Phi(x^*) \leq \Phi(\bar{x})$.

Consider minimising

$$\min_{x \in \mathbb{R}^n} \{ \Phi(x) \stackrel{\text{def}}{=} R(x) + F(x) \},$$

- $R : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ is proper l.s.c. and bounded from below.
- $F : \mathbb{R}^n \rightarrow \mathbb{R}$ is finite-valued, differentiable and ∇F is L -Lipschitz.

Forward-Backward splitting

Let $\gamma \in]0, 1/L[$:

$$x_{k+1} \in \text{prox}_{\gamma R}(x_k - \gamma \nabla F(x_k)).$$

- **Sufficient decrease**

$$\Phi(x_{k+1}) + \frac{1-\gamma L}{2\gamma} \|x_k - x_{k+1}\|^2 \leq \Phi(x_k).$$

- **Relative error** $g_{k+1} \stackrel{\text{def}}{=} \frac{1}{\gamma}(x_k - x_{k+1}) - \nabla F(x_k) + \nabla F(x_{k+1}) \in \partial \Phi(x_{k+1})$

$$\|g_{k+1}\| \leq \left(\frac{1}{\gamma} + L \right) \|x_k - x_{k+1}\|.$$

- **Continuity** sequence $\{x_k\}_{k \in \mathbb{N}}$ is bounded.

A coupled problem

Consider minimising

$$\min_{x \in \mathbb{R}^n, y \in \mathbb{R}^m} \{E(x, y) \stackrel{\text{def}}{=} R(x) + F(x, y) + J(y)\},$$

- $R : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}, J : \mathbb{R}^m \rightarrow \mathbb{R} \cup \{+\infty\}$ are proper l.s.c. and bounded from below.
- $F : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ is finite-valued, differentiable and ∇F is L -Lipschitz.

Subdifferential

$$\partial E(x, y) = \{\partial R(x) + \nabla_x F(x, y)\} \times \{\partial J(y) + \nabla_y F(x, y)\} = \partial_x E(x, y) \times \partial_y E(x, y).$$

Separate Lipschitz continuity for F : $\nabla_x F$ is L_x -Lip. and $\nabla_y F$ is L_y -Lip.

PAM is an alternating minimisation algorithm.

PAM

Let $\gamma_x, \gamma_y \in]0, 1/L[$:

$$x_{k+1} \in \operatorname{argmin}_{x \in \mathbb{R}^n} E(x, y_k) + \frac{1}{2\gamma_x} \|x - x_k\|^2,$$

$$y_{k+1} \in \operatorname{argmin}_{y \in \mathbb{R}^m} E(x_{k+1}, y) + \frac{1}{2\gamma_y} \|y - y_k\|^2.$$

- PAM is an instance of PPA.
- Convergence, let $\Phi(x, y) = E(x, y)$.
- No closed form solution,

$$\begin{aligned} x_{k+1} &\in \operatorname{argmin}_{x \in \mathbb{R}^n} E(x, y_k) + \frac{1}{2\gamma_x} \|x - x_k\|^2 \\ &= \operatorname{argmin}_{x \in \mathbb{R}^n} R(x) + F(x, y_k) + \frac{1}{2\gamma_x} \|x - x_k\|^2. \end{aligned}$$

PALM is linearised PAM.

$$F(x, y_k) \leq F(x_k, y_k) + \langle \nabla_x F(x_k, y_k), x - x_k \rangle + \frac{1}{2\gamma_x} \|x - x_k\|^2.$$

PALM

Let $\gamma_x, \gamma_y \in]0, 1/L[$:

$$x_{k+1} \in \text{prox}_{\gamma_x R}(x_k - \gamma_x \nabla_x F(x_k, y_k)),$$

$$y_{k+1} \in \text{prox}_{\gamma_y J}(y_k - \gamma_y \nabla_y F(x_{k+1}, y_k)).$$

- PAM is an instance of Forward-Backward.
- Convergence, let $\Phi(x, y) = E(x, y)$.

- Converges to global minimiser if starts close enough.
- Inertial acceleration can be applied to all of them.
- Step-size v.s. inertial parameter.
- Step-size and critical points.
- Stochastic optimisation methods can escape saddle-point or find global minimiser...

- H. Attouch, and J. Bolte. “On the convergence of the proximal algorithm for nonsmooth functions involving analytic features”. *Mathematical Programming* 116.1-2 (2009): 5-16.
- H. Attouch, J. Bolte, and B. Svaiter. “Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods”. *Mathematical Programming* 137.1-2 (2013): 91-129.
- H. Attouch, et al. “Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Łojasiewicz inequality”. *Mathematics of Operations Research* 35.2 (2010): 438-457.
- J. Bolte, S. Sabach, and M. Teboulle. “Proximal alternating linearized minimization or nonconvex and nonsmooth problems”. *Mathematical Programming* 146.1-2 (2014): 459-494.
- J. Liang, J. Fadili, and G. Peyré. “A multi-step inertial forward-backward splitting method for non-convex optimization”. *Advances in Neural Information Processing Systems*. 2016.