UW Applied Mathematics
The Next 50 Years

Introduction to Convex–Composite Optimization

Jim Burke

Outline

1. Convexity
   i. Sets and functions
   ii. Supporting hyperplanes and support functions
   iii. Convex Conjugates and subgradients
   iv. Bi-conjugacy and subgradients

2. Convex-Composite Optimization
   i. Problem statement, history, and examples
   ii. Piecewise linear quadratic penalties
   iii. The convex composite Lagrangian and optimality conditions
   iv. Structure of algorithms
   v. Quadratic convergence of Newton's method
   vi. Globalization techniques

Convexity

Convex Sets: A subset C of $\mathbb{R}^n$ is convex if

$$[x, y] \subset C \quad \forall\, x, y \in C,$$

where $[x, y] := \{(1 - \lambda x) + \lambda y \mid 0 \leq \lambda \leq 1\}$ is the line segment connecting x and y.

Convexity

Convex Sets: A subset C of $\mathbb{R}^n$ is convex if

$$[x, y] \subset C \quad \forall\, x, y \in C,$$

where $[x, y] := \{(1 - \lambda x) + \lambda y \mid 0 \leq \lambda \leq 1\}$ is the line segment connecting x and y.

Convex Functions: $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is said to be convex if

$$\operatorname{epi} f := \{(x, \mu) \mid f(x) \leq \mu\},$$

is a convex set.

Convexity

Convex Sets: A subset C of $\mathbb{R}^n$ is convex if

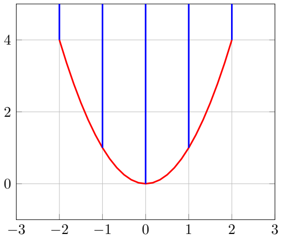
$$[x, y] \subset C \quad \forall\, x, y \in C,$$

where $[x, y] := \{(1 - \lambda x) + \lambda y \mid 0 \leq \lambda \leq 1\}$ is the line segment connecting x and y.

Convex Functions: $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is said to be convex if

$$\operatorname{epi} f := \{(x, \mu) \mid f(x) \leq \mu\},$$

is a convex set.



Convexity

Convex Sets: A subset C of $\mathbb{R}^n$ is convex if

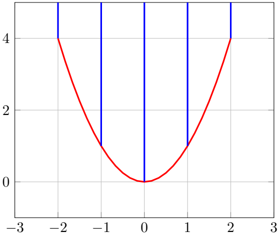
$$[x, y] \subset C \quad \forall\, x, y \in C,$$

where $[x, y] := \{(1-\lambda x) + \lambda y \mid 0 \le \lambda \le 1\}$ is the line segment connecting x and y.

Convex Functions: $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is said to be convex if

$$\operatorname{epi} f := \{(x, \mu) \mid f(x) \le \mu\},$$

is a convex set.



f is lower semi-continuous (lsc) $\iff$ epi (f) is closed

Examples

Sets:

- Subspaces and affine sets (shifted subspaces).
- Hyperplanes: affine sets of co-dimension 1.
- The unit ball of any norm.
- Convex cones: $K \subset \mathbb{R}^n$ is a convex cone if
$$\lambda K \subset K \ \ \forall \lambda > 0 \ \text{ and } K + K \subset K,$$
e.g. $\mathbb{R}^n_+$ and $\mathbb{S}^n{}_+$.

Examples

Sets:

- Subspaces and affine sets (shifted subspaces).
- Hyperplanes: affine sets of co-dimension 1.
- The unit ball of any norm.
- Convex cones: $K \subset \mathbb{R}^n$ is a convex cone if
$$\lambda K \subset K \ \ \forall \lambda > 0 \ \text{ and } K + K \subset K,$$
e.g. $\mathbb{R}^n_+$ and $\mathbb{S}^n{}_+$.

Functions:

- Linear functionals, $x \mapsto \langle z, x \rangle$.
- Any norm, and a norm to a power greater than 1.
- The exponential function and the negative of a logarithm.
- Indicators function of convex sets: $C \subset \mathbb{R}^n$ convex,
$$\delta_C(x) := \begin{cases} 0 & , \ x \in C, \\ +\infty & , \ x \notin C. \end{cases}$$
- Support function of convex sets: $C \subset \mathbb{R}^n$ convex,
$$\sigma_C(y) := \sup \{ \langle y, x \rangle \mid x \in C \}.$$

Relative Interiors and Supporting Hyperplanes

Relative interior: The relative interior of a convex set C is the interior relative to the smallest affine set that contains C, $\mathrm{aff}\,(C)$:

$$\mathrm{ri}\, C := \{x \in C \mid \exists \epsilon > 0 \text{ s.t. } (x + \epsilon \mathbb{B}) \cap \mathrm{aff}\,(C) \subset C\}$$

Relative Interiors and Supporting Hyperplanes

Relative interior: The relative interior of a convex set C is the interior relative to the smallest affine set that contains C, $\mathrm{aff}\,(C)$:

$$\mathrm{ri}\, C := \{x \in C \mid \exists \epsilon > 0 \text{ s.t. } (x + \epsilon\mathbb{B}) \cap \mathrm{aff}\,(C) \subset C\}$$

THM: If $\overline{x} \in \mathrm{rbdry}\,(C) := \mathrm{cl}\, C \setminus \mathrm{ri}\, C$, then there is a hyperplane $H_z := \{x \mid \langle z, x\rangle = \langle z, \overline{x}\rangle\}$ such that $H_z \cap C = \emptyset$, i.e.,

$$\exists\, z \text{ s.t. } \langle z, x\rangle < \langle z, \overline{x}\rangle \quad \forall\, x \in \mathrm{ri}\, C.$$

Relative Interiors and Supporting Hyperplanes

Relative interior: The relative interior of a convex set C is the interior relative to the smallest affine set that contains C, $\mathrm{aff}\,(C)$:

$$\mathrm{ri}\, C := \{x \in C \mid \exists \epsilon > 0 \text{ s.t. } (x + \epsilon\mathbb{B}) \cap \mathrm{aff}\,(C) \subset C\}$$

THM: If $\overline{x} \in \mathrm{rbdry}\,(C) := \mathrm{cl}\, C \setminus \mathrm{ri}\, C$, then there is a hyperplane $H_z := \{x \mid \langle z, x\rangle = \langle z, \overline{x}\rangle\}$ such that $H_z \cap C = \emptyset$, i.e.,

$$\exists\, z \text{ s.t. } \langle z, x\rangle < \langle z, \overline{x}\rangle \quad \forall\, x \in \mathrm{ri}\, C.$$

H_z is said to be a supporting hyperplane to C at $\overline{x}$ with support vector z.

Relative Interiors and Supporting Hyperplanes

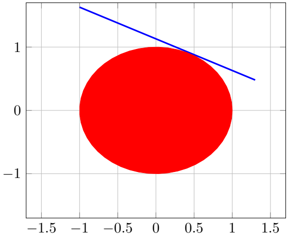
Relative interior: The relative interior of a convex set C is the interior relative to the smallest affine set that contains C, $\mathrm{aff}\,(C)$:

$$\mathrm{ri}\, C := \{x \in C \mid \exists \epsilon > 0 \text{ s.t. } (x + \epsilon\mathbb{B}) \cap \mathrm{aff}\,(C) \subset C\}$$

THM: If $\overline{x} \in \mathrm{rbdry}\,(C) := \mathrm{cl}\, C \setminus \mathrm{ri}\, C$, then there is a hyperplane $H_z := \{x \mid \langle z, x\rangle = \langle z, \overline{x}\rangle\}$ such that $H_z \cap C = \emptyset$, i.e.,

$$\exists\, z \text{ s.t. } \langle z, x\rangle < \langle z, \overline{x}\rangle \quad \forall\, x \in \mathrm{ri}\, C.$$

H_z is said to be a supporting hyperplane to C at $\overline{x}$ with support vector z.



Support functions

For a closed convex $C \subset \mathbb{R}^n$,

$$\overline{x} \in \operatorname{argmax}_C \langle z, x \rangle \quad \Longleftrightarrow \quad z \text{ supports } C \text{ at } \overline{x}.$$

Support functions

For a closed convex $C \subset \mathbb{R}^n$,

$$\overline{x} \in \operatorname{argmax}_C \langle z, x \rangle \iff z \text{ supports } C \text{ at } \overline{x}.$$

Fact: Support functions are sublinear:

Support functions

For a closed convex $C \subset \mathbb{R}^n$,

$$\bar{x} \in \operatorname{argmax}_C \langle z, x \rangle \iff z \text{ supports } C \text{ at } \bar{x}.$$

Fact: Support functions are sublinear:

1. (positively homogeneous) $\sigma(\lambda x) = \lambda \sigma(x) \; \forall \, \lambda \geq 0$,

Support functions

For a closed convex $C \subset \mathbb{R}^n$,

$$\overline{x} \in \operatorname{argmax}_C \langle z, x \rangle \iff z \text{ supports } C \text{ at } \overline{x}.$$

Fact: Support functions are sublinear:

1. (positively homogeneous) $\sigma(\lambda x) = \lambda\sigma(x) \; \forall \, \lambda \geq 0$,
2. (subadditive) $\sigma(x + y) \leq \sigma(x) + \sigma(y)$.

Support functions

For a closed convex $C \subset \mathbb{R}^n$,

$$\bar{x} \in \operatorname{argmax}_C \langle z, x \rangle \iff z \text{ supports } C \text{ at } \bar{x}.$$

Fact: Support functions are sublinear:

1. (positively homogeneous) $\sigma(\lambda x) = \lambda \sigma(x) \; \forall \, \lambda \geq 0$,
2. (subadditive) $\sigma(x + y) \leq \sigma(x) + \sigma(y)$.

Hörmander's Theorem: $\sigma : \mathbb{R}^n \to \bar{\mathbb{R}}_+ := \mathbb{R}_+ \cup \{+\infty\}$ lsc.

σ is sublinear $\iff$ epi (σ) is a closed cvx cone $\iff$ $\sigma = \sigma_C$,

where $C := \{z \mid \langle z, x \rangle \leq f(x) \; \forall \, x\} = \{z \mid \langle z, x \rangle \leq 1 \; \forall \, f(x) \leq 1\}$.

The Convex Conjugate and Subgradients

$f : \mathbb{R}^n \to \mathbb{R} \cup +\infty$ convex, i.e., epi f is convex.

$$\begin{aligned} \sigma_{\operatorname{epi} f}\left((z,-1)\right) &= \sup_{f(x)\le\mu} \langle (z,-1),(x,\mu)\rangle \\ &= \sup_x [\langle z,x\rangle - f(x)] \qquad =: f^*(z) \end{aligned}$$

The Convex Conjugate and Subgradients

$f : \mathbb{R}^n \to \mathbb{R} \cup +\infty$ convex, i.e., epi f is convex.

$$\begin{aligned}\sigma_{\mathrm{epi}\, f}((z,-1)) &= \sup_{f(x)\le\mu} \langle (z,-1),(x,\mu)\rangle \\ &= \sup_x [\langle z,x\rangle - f(x)] \qquad =: f^*(z)\end{aligned}$$

Subgradients:

$$\bar{x} \in \operatorname*{argmax}_x [\langle z,x\rangle - f(x)] \iff (z,-1) \text{ supports epi}\,(f) \text{ at } (\bar{x}, f(\bar{x})).$$

The Convex Conjugate and Subgradients

$f : \mathbb{R}^n \to \mathbb{R} \cup +\infty$ convex, i.e., epi f is convex.

$$\begin{aligned}\sigma_{\text{epi}\, f}\left((z,-1)\right) &= \sup_{f(x)\le\mu} \langle (z,-1),(x,\mu)\rangle \\ &= \sup_x [\langle z,x\rangle - f(x)] \qquad =: f^*(z)\end{aligned}$$

Subgradients:

$$\overline{x} \in \operatorname*{argmax}_x [\langle z,x\rangle - f(x)] \iff (z,-1) \text{ supports epi}\,(f) \text{ at } (\overline{x}, f(\overline{x})).$$

$$\iff \langle (z,-1),(\overline{x}, f(\overline{x})\rangle \geq \langle (z,-1),(x,f(x))\rangle \ \forall\, x \in \operatorname{dom} f,$$

The Convex Conjugate and Subgradients

$f : \mathbb{R}^n \to \mathbb{R} \cup +\infty$ convex, i.e., epi f is convex.

$$\sigma_{\text{epi } f}((z,-1)) = \sup_{f(x)\le\mu} \langle (z,-1),(x,\mu)\rangle$$
$$= \sup_x [\langle z,x\rangle - f(x)] \qquad =: f^*(z)$$

Subgradients:

$$\bar{x} \in \operatorname*{argmax}_x [\langle z,x\rangle - f(x)] \iff (z,-1) \text{ supports epi}\,(f) \text{ at } (\bar{x}, f(\bar{x})).$$

$$\iff \langle (z,-1),(\bar{x}, f(\bar{x})\rangle \ge \langle (z,-1),(x,f(x))\rangle \ \forall\, x \in \operatorname{dom} f,$$

$$\iff f(x) \ge f(\bar{x}) + \langle z, x-\bar{x}\rangle \quad \forall\, x \in \mathbb{R}^n$$

The Convex Conjugate and Subgradients

$f\ :\ \mathbb{R}^n \to \mathbb{R} \cup +\infty$ convex, i.e., epi f is convex.

$$\begin{aligned}\sigma_{\mathrm{epi}\,f}\left((z,-1)\right) &= \sup_{f(x)\le\mu} \langle (z,-1),(x,\mu)\rangle \\ &= \sup_x [\langle z,x\rangle - f(x)] \qquad =: f^*(z)\end{aligned}$$

Subgradients:

$$\overline{x} \in \operatorname*{argmax}_x [\langle z,x\rangle - f(x)] \iff (z,-1) \text{ supports } \mathrm{epi}\,(f) \text{ at } (\overline{x}, f(\overline{x})).$$

$$\iff \langle (z,-1),(\overline{x}, f(\overline{x})\rangle \ge \langle (z,-1),(x,f(x))\rangle \ \forall\, x \in \mathrm{dom}\, f,$$

$$\iff f(x) \ge f(\overline{x}) + \langle z, x-\overline{x}\rangle \quad \forall\, x \in \mathbb{R}^n$$

$\partial f(\overline{x}) := \operatorname*{argmax}_x [\langle z,x\rangle - f(x)]$, the subdifferential of f at $\overline{x}$,

The Convex Conjugate and Subgradients

$f : \mathbb{R}^n \to \mathbb{R} \cup +\infty$ convex, i.e., epi f is convex.

$$\sigma_{\mathrm{epi}\, f}\left((z, -1)\right) = \sup_{f(x) \leq \mu} \langle (z, -1), (x, \mu) \rangle$$
$$= \sup_x [\langle z, x \rangle - f(x)] \qquad =: f^*(z)$$

Subgradients:

$$\overline{x} \in \operatorname*{argmax}_x [\langle z, x \rangle - f(x)] \iff (z, -1) \text{ supports epi}\,(f) \text{ at } (\overline{x}, f(\overline{x})).$$

$$\iff \langle (z, -1), (\overline{x}, f(\overline{x}) \rangle \geq \langle (z, -1), (x, f(x)) \rangle \;\forall\, x \in \mathrm{dom}\, f,$$

$$\iff f(x) \geq f(\overline{x}) + \langle z, x - \overline{x} \rangle \quad \forall\, x \in \mathbb{R}^n$$

$\partial f(\overline{x}) := \operatorname*{argmax}_x [\langle z, x \rangle - f(x)]$, the subdifferential of f at $\overline{x}$,

$\partial f(\overline{x})$ a non-empty closed convex set on $\mathrm{ri}\,\mathrm{dom}\, f$.

Bi-Conjugacy and Subgradients

$$f^*(z) := \sup_x[\langle z, x\rangle - f(x)]$$

$$\begin{aligned} f^*(z) \geq \langle z, x\rangle - f(x) \quad &\forall \;\; x \in \operatorname{dom}(f) \text{ and } z \in \mathbb{R}^n \\ \iff \quad f(x) \geq \langle z, x\rangle - f^*(z) \qquad &\forall \;\; z \in \operatorname{dom}(f^*) \text{ and } x \in \mathbb{R}^n \\ \implies \quad f(x) \geq f^{**}(x) \qquad &\forall\, x \in \mathbb{R}^n \end{aligned}$$

Bi-Conjugacy and Subgradients

$$f^*(z) := \sup_x[\langle z, x\rangle - f(x)]$$

$$\begin{aligned} f^*(z) \geq \langle z, x\rangle - f(x) \quad & \forall \;\; x \in \operatorname{dom}(f) \text{ and } z \in \mathbb{R}^n \\ \iff \quad f(x) \geq \langle z, x\rangle - f^*(z) \quad & \forall \;\; z \in \operatorname{dom}(f^*) \text{ and } x \in \mathbb{R}^n \\ \implies \quad f(x) \geq f^{**}(x) \quad & \forall\, x \in \mathbb{R}^n \end{aligned}$$

But

$$z \in \partial f(x) \quad \iff \quad \langle z, x\rangle \geq f(x) + f^*(z),$$

so $\forall\, x \in \operatorname{dom}(\partial f) := \{x \mid \partial f(x) \neq \emptyset\}$ and $z \in \partial f(x)$,

$$f(x) \leq \langle z, x\rangle - f^*(z) \leq \sup_w[\langle w, x\rangle - f^*(w)] = f^{**}(x) \leq f(x).$$

Bi-Conjugacy and Subgradients

$$f^*(z) := \sup_x[\langle z, x\rangle - f(x)]$$

$$\begin{aligned} f^*(z) \geq \langle z, x\rangle - f(x) \quad & \forall \;\; x \in \operatorname{dom}(f) \text{ and } z \in \mathbb{R}^n \\ \iff \quad f(x) \geq \langle z, x\rangle - f^*(z) \qquad & \forall \;\; z \in \operatorname{dom}(f^*) \text{ and } x \in \mathbb{R}^n \\ \implies \quad f(x) \geq f^{**}(x) \qquad & \forall\, x \in \mathbb{R}^n \end{aligned}$$

But

$$z \in \partial f(x) \quad \iff \quad \langle z, x\rangle \geq f(x) + f^*(z),$$

so $\forall\, x \in \operatorname{dom}(\partial f) := \{x \,|\, \partial f(x) \neq \emptyset\}$ and $z \in \partial f(x)$,

$$f(x) \leq \langle z, x\rangle - f^*(z) \leq \sup_w[\langle w, x\rangle - f^*(w)] = f^{**}(x) \leq f(x).$$

So $f(x) = f^{**}(x)$ on $\operatorname{dom}(\partial f)$, where $\operatorname{ri}\operatorname{dom}(f) \subset \operatorname{dom}(\partial f)$.
Consequently $f^{**} = \operatorname{cl} f$, and if $f = \operatorname{cl} f$, then

$$f = f^{**} \text{ and } \partial f^* = (\partial f)^{-1}.$$

The Subdifferential and the Directional Derivative

$$z \in \partial f(x) \iff f(y) \geq f(x) + \langle z, y - x \rangle \qquad \forall\ y \in \mathbb{R}^n$$

The Subdifferential and the Directional Derivative

$$z \in \partial f(x) \iff f(y) \geq f(x) + \langle z, y - x \rangle \qquad \forall\, y \in \mathbb{R}^n$$

Hence,

$$\phi(t) := \frac{f(x + td) - f(x)}{t} \geq \langle z, d \rangle \qquad \forall\, d \in \mathbb{R}^n \text{ and } t > 0.$$

The Subdifferential and the Directional Derivative

$$z \in \partial f(x) \iff f(y) \geq f(x) + \langle z, y - x \rangle \qquad \forall\, y \in \mathbb{R}^n$$

Hence,

$$\phi(t) := \frac{f(x + td) - f(x)}{t} \geq \langle z, d \rangle \qquad \forall\, d \in \mathbb{R}^n \text{ and } t > 0.$$

It is easily seen that ϕ is nondecreasing for $t > 0$, hence,

$$f'(x; d) := \lim_{t \downarrow 0} \frac{f(x + td) - f(x)}{t} = \inf_{t > 0} \frac{f(x + td) - f(x)}{t} \geq \sigma_{\partial f(x)}(d).$$

The Subdifferential and the Directional Derivative

$$z \in \partial f(x) \iff f(y) \geq f(x) + \langle z, y - x \rangle \qquad \forall\, y \in \mathbb{R}^n$$

Hence,

$$\phi(t) := \frac{f(x+td) - f(x)}{t} \geq \langle z, d \rangle \qquad \forall\, d \in \mathbb{R}^n \text{ and } t > 0.$$

It is easily seen that ϕ is nondecreasing for $t > 0$, hence,

$$f'(x; d) := \lim_{t \downarrow 0} \frac{f(x+td) - f(x)}{t} = \inf_{t>0} \frac{f(x+td) - f(x)}{t} \geq \sigma_{\partial f(x)}(d).$$

Moreover, $f'(x; d)$ is easily seen to be sublinear from which one can show that

$$f'(x; d) = \sigma_{\partial f(x)}(d).$$

Convex-Composite Optimization (Non-Convex)

$$\min_{x \in \mathbb{R}^n} f(x) := h(c(x)) \qquad \textbf{(P)}$$

$h : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ is closed, proper, convex
$c : \mathbb{R}^n \to \mathbb{R}^m$ is $\mathcal{C}^2$-smooth

Convex-Composite Optimization (Non-Convex)

$$\min_{x \in \mathbb{R}^n} f(x) := h(c(x)) \qquad (\mathbf{P})$$

$h : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ is closed, proper, convex — The Model

$c : \mathbb{R}^n \to \mathbb{R}^m$ is $\mathcal{C}^2$-smooth — The Data

Convex-Composite Optimization (Non-Convex)

$$\min_{x\in\mathbb{R}^n} f(x) := h(c(x)) \quad + g(x) \qquad (\mathbf{P})$$

$h : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ is closed, proper, convex — The Model

$c : \mathbb{R}^n \to \mathbb{R}^m$ is $\mathcal{C}^2$-smooth — The Data

$g : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is closed, proper, convex — Regularizer

used to induce solution properties

Convex-Composite Optimization (Non-Convex)

$$\min_{x \in \mathbb{R}^n} f(x) := h(c(x)) \quad + g(x) \qquad \textbf{(P)}$$

$h : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ is closed, proper, convex
$c : \mathbb{R}^n \to \mathbb{R}^m$ is $\mathcal{C}^2$-smooth
$g : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is closed, proper, convex   Regularizer
used to induce solution properties

70's
Fletcher, Powel, Osborne

80-90's
Burke, Ferris, Fletcher, Kawasaki, Masden, Poliquin, Powel, Osborne, Rockafellar, Womersley, Wright, Yuan

Recent (15-19's)
Aravkin, Bell, B, Chang, Cui, Duchi, Davis, Drusvyatskiy, Hoheisel, Hong, Lewis, Ioffe, Mordukhovich, Pang, Ruan

Examples:

Non-linear least-squares: $f(x) = \|c(x)\|_2^2$

Examples:

Non-linear least-squares: $f(x) = \|c(x)\|_2^2$

Feasibility Problems: $c(x) \in C: \quad \min \operatorname{dist}(c(x) \,|\, C),$

where $C \subset \mathbb{R}^m$ is non-empty, closed, convex, and
$\operatorname{dist}(y \,|\, C) := \inf \{ \|y - z\| \} \, z \in C.$

Examples:

Non-linear least-squares: $f(x) = \|c(x)\|_2^2$

Feasibility Problems: $c(x) \in C: \quad \min \operatorname{dist}(c(x) \,|\, C),$

where $C \subset \mathbb{R}^m$ is non-empty, closed, convex, and

$\operatorname{dist}(y \,|\, C) := \inf \{ \|y - z\| \} \, z \in C.$

Exact Penalization: $\min \varphi(x) + \alpha \operatorname{dist}(\hat{c}(x) \,|\, C)$

Here $c(x) := (\varphi(x), \hat{c}(x))$ and $h(\mu, y) := \mu + \alpha \operatorname{dist}(y \,|\, C)$

Examples:

Non-linear least-squares: $f(x) = \|c(x)\|_2^2$

Feasibility Problems: $c(x) \in C$: $\quad \min \operatorname{dist}(c(x) \,|\, C)$,
where $C \subset \mathbb{R}^m$ is non-empty, closed, convex, and
$\operatorname{dist}(y \,|\, C) := \inf \{ \|y - z\| \} \, z \in C$.

Exact Penalization: $\min \varphi(x) + \alpha \operatorname{dist}(\hat{c}(x) \,|\, C)$
Here $c(x) := (\varphi(x), \hat{c}(x))$ and $h(\mu, y) := \mu + \alpha \operatorname{dist}(y \,|\, C)$

Non-linear programming: $\min \varphi(x) + \delta_C(\hat{c}(x))$.
Here $c(x) := (\varphi(x), \hat{c}(x))$ and $h(\mu, y) := \mu + \delta_C(y)$, where
$\delta_C(y) = 0$ if $y \in C$ and $+\infty$ otherwise.

More Recent Examples

Optimal Value Composition:

$$h(c) := \min \{ b^\top y \} \, Ay \leq c$$

More Recent Examples

Optimal Value Composition:

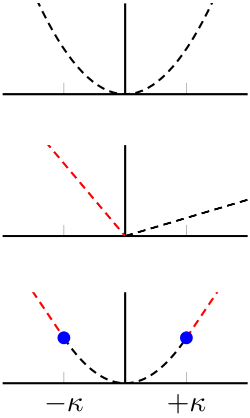
$$h(c) := \min \{ b^\top y \} \, Ay \leq c$$

Piecewise linear-quadratic (PLQ) penalties:
(Rockfellar-Wets (97))

$$h(c) := \sup_{u \in U} \langle u, Bc \rangle - \frac{1}{2} u^T M u$$

with $U \subset \mathbb{R}^k$ non-empty, polyhedral, closed, convex, $M \in \mathbb{S}^n$ is positive semi-definite.

Dual representation of PLQ Penalties

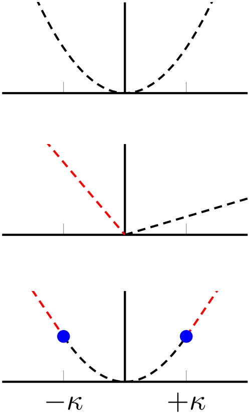


$$\frac{1}{2}x^2 = \sup_{u\in\mathbb{R}} \langle u, x\rangle - \frac{1}{2}u^2$$

$$Q_{0.8}(x) = \sup_{u\in[-0.8,0.2]} \langle u, x\rangle$$

$$\rho_h(x) = \sup_{u\in[-\kappa,\kappa]} \langle u, x\rangle - \frac{1}{2}u^2$$

Dual representation of PLQ Penalties



$$\frac{1}{2}x^2 = \sup_{u \in \mathbb{R}} \langle u, x\rangle - \frac{1}{2}u^2$$

$$Q_{0.8}(x) = \sup_{u \in [-0.8, 0.2]} \langle u, x\rangle$$

$$\rho_h(x) = \sup_{u \in [-\kappa, \kappa]} \langle u, x\rangle - \frac{1}{2}u^2$$

PLQ penalties closed under addition and affine composition.

PLQ penalties in practice

Application	Objective	PLQs
Regression	$\|Ax - b\|^2$	L_2
Robust regression	$\rho_H(Ax - b)$	Huber
Quantile regression	$Q(Ax - b)$	Asym. L_1
Lasso	$\|Ax - b\|^2 + \lambda\|x\|_1$	$L_2 + L_1$
Robust lasso	$\rho_H(Ax - b) + \lambda\|x\|_1$	Huber + L_1
SVM	$\frac{1}{2}\|w\|^2 + H(\mathbf{1} - Ax)$	L_1 + hinge loss
SVR	$\rho_V(Ax - b)$	Vapnik loss
Kalman smoother	$\|Gx - w\|^2_{Q^{-1}} + \|Hx - z\|^2_{R^{-1}}$	$L_2 + L_2$
Robust trend smoothing	$\|Gx - w\|_1 + \rho_H(Hx - z)$	L_1 + Huber

The Convex-Composite Lagrangian

$$\mathbf{P} \qquad \min_{x \in \mathbb{R}^n} h(c(x))$$

- The Lagrangian for **P**: (B. (87))

$$L(x, y) := \langle y, c(x) \rangle - h^*(y)$$

The Convex-Composite Lagrangian

$$\mathbf{P} \qquad \min_{x\in\mathbb{R}^n} h(c(x)) \quad + g(x)$$

- The Lagrangian for **P**: (B. (87))

$$L(x,y) := \langle y, c(x)\rangle - h^*(y) \quad + g(x)$$

The Convex-Composite Lagrangian

$$\mathbf{P} \qquad \min_{x\in\mathbb{R}^n} h(c(x)) \quad + g(x)$$

- The Lagrangian for **P**: (B. (87))

$$L(x, y, v) := \langle y, c(x)\rangle - h^*(y) \quad + \langle v, x\rangle - g^*(v)$$

The Convex-Composite Lagrangian

$$\mathbf{P} \qquad \min_{x\in\mathbb{R}^n} h(c(x))$$

- The Lagrangian for **P**: (B. (87))

$$L(x,y) := \langle y, c(x)\rangle - h^*(y) \qquad \begin{cases} \text{(Primal)} & \inf_x \sup_y \; L(x,y) \\ & \\ \text{(Dual)} & \sup_y \inf_x \; L(x,y) \end{cases}$$

The Convex-Composite Lagrangian

$$\mathbf{P} \qquad \min_{x \in \mathbb{R}^n} h(c(x))$$

- The Lagrangian for **P**: (B. (87))

$$L(x,y) := \langle y, c(x)\rangle - h^*(y) \qquad \begin{cases} \text{(Primal)} & \inf_x \sup_y \; L(x,y) \\ \\ \text{(Dual)} & \sup_y \inf_x \; L(x,y) \end{cases}$$

- The basic constraint qualification (BCQ): Rockafellar (88)

$$\ker c'(x) \cap N\left(c(x) \,|\, \operatorname{dom} h\right) = \{0\}$$

The Convex-Composite Lagrangian

$$\mathbf{P} \qquad \min_{x\in\mathbb{R}^n} h(c(x))$$

- The Lagrangian for **P**: (B. (87))

$$L(x,y) := \langle y, c(x)\rangle - h^*(y) \qquad \begin{cases} \text{(Primal)} & \inf_x \sup_y \; L(x,y) \\ \\ \text{(Dual)} & \sup_y \inf_x \; L(x,y) \end{cases}$$

- The basic constraint qualification (BCQ): Rockafellar (88)

$$\ker c'(x) \cap N\left(c(x)\,|\,\mathrm{dom}\, h\right) = \{0\}$$

- The subdifferential under BCQ: $\partial f(x) := c'(x)^T \partial h(c(x))$.

The Convex-Composite Lagrangian

$$\mathbf{P} \qquad \min_{x\in\mathbb{R}^n} h(c(x))$$

- The Lagrangian for **P**: (B. (87))

$$L(x,y) := \langle y, c(x)\rangle - h^*(y) \qquad \begin{cases} \text{(Primal)} & \inf_x \sup_y \; L(x,y) \\ & \\ \text{(Dual)} & \sup_y \inf_x \; L(x,y) \end{cases}$$

- The basic constraint qualification (BCQ): Rockafellar (88)

$$\ker c'(x) \cap N\left(c(x)\,|\,\operatorname{dom} h\right) = \{0\}$$

- The subdifferential under BCQ: $\partial f(x) := c'(x)^T \partial h(c(x))$.
- First-Order Optimality Conditions:

$$[\min_x f \quad \mapsto \overline{x}] \implies 0 \in \partial f(\overline{x}) \iff \begin{pmatrix} 0 \\ 0 \end{pmatrix} \in \begin{pmatrix} \partial_x L(\overline{x},\overline{y}) \\ \partial_y(-L)(\overline{x},\overline{y}) \end{pmatrix}$$

Algorithms

$$\mathbf{P}_k \qquad \min_{\|x-x^k\|\le\eta_k} h\left(c(x^k)+\nabla c(x^k)[x-x^k]\right)+\frac{1}{2}(x-x^k)^\top H_k(x-x^k),$$

- H_k approximates the Hessian of a Lagrangian for $\mathbf{P}$ at (x^k, y^k)
- Newton's method: $H_k := \nabla^2_{xx}L(x^k, y^k) = \sum_{k=1}^m y_i^k \nabla^2_{xx} c_i(x^k)$ and $\eta_k \equiv +\infty$
- $\mathbf{P}_k$ may or may not be convex depending on whether $H_k \succeq 0$.
- A example is the Gauss-Newton method: $h = \|\cdot\|_2^2$

$$\min_x \left\|c(x^k) + c'(x^k)(x - x^k)\right\|_2^2$$

Algorithm for NLP

NLP minimize $\phi(x)$

subject to $f_i(x) = 0, \; i = 1, \ldots, s, \; f_i(x) \leq 0, \; i = s+1, \ldots, m.$

Algorithm for NLP

$$\begin{array}{lll} \text{NLP} & \text{minimize} & \phi(x) \\ & \text{subject to} & f_i(x)=0,\ i=1,\ldots,s,\ f_i(x)\le 0,\ i=s+1,\ldots,m. \end{array}$$

- Convex-Composite Framework

$$\begin{aligned} h(\mu,y) &= \mu + \delta_K(y), && K := \{0\}^s \times \mathbb{R}^{m-s}_- \\ c(x) &= (\phi(x), f(x)) && \\ L(x,y) &= \phi(x) + \sum_{k=1}^m y_i f_i(x) \ - \delta_{K^\circ}(y), && K^\circ = \mathbb{R}^s \times \mathbb{R}^{m-s}_+ \end{aligned}$$

Algorithm for NLP

$$\begin{array}{lll} \text{NLP} & \text{minimize} & \phi(x) \\ & \text{subject to} & f_i(x)=0,\ i=1,\ldots,s,\ f_i(x)\le 0,\ i=s+1,\ldots,m. \end{array}$$

- Convex-Composite Framework

$$\begin{aligned} h(\mu,y) &= \mu + \delta_K(y), & K &:= \{0\}^s \times \mathbb{R}^{m-s}_- \\ c(x) &= (\phi(x), f(x)) \\ L(x,y) &= \phi(x) + \sum_{k=1}^m y_i f_i(x)\ -\delta_{K^\circ}(y), & K^\circ &= \mathbb{R}^s \times \mathbb{R}^{m-s}_+ \end{aligned}$$

- Subproblems:

$$\begin{array}{lll} \mathbf{P_k} & \text{minimize} & \phi(x^k) + \nabla\phi(x^k)^T(x-x^k) + \frac{1}{2}[x-x^k]^\top H_k[x-x^k] \\ & \text{subject to} & f_i(x^k) + \nabla f_i(x^k)^T(x-x^k) = 0,\ i=1,\ldots,s \\ & & f_i(x^k) + \nabla f_i(x^k)^T(x-x^k) = 0,\ i=s+1,\ldots,m. \end{array}$$

Algorithm for NLP

$$\begin{array}{lll} \text{NLP} & \text{minimize} & \phi(x) \\ & \text{subject to} & f_i(x)=0,\ i=1,\dots,s,\ f_i(x)\le 0,\ i=s+1,\dots,m. \end{array}$$

- Convex-Composite Framework

$$\begin{aligned} h(\mu, y) &= \mu + \delta_K(y), & K &:= \{0\}^s \times \mathbb{R}^{m-s}_- \\ c(x) &= (\phi(x), f(x)) \\ L(x,y) &= \phi(x) + \sum_{k=1}^{m} y_i f_i(x) \ - \delta_{K^\circ}(y), & K^\circ &= \mathbb{R}^s \times \mathbb{R}^{m-s}_+ \end{aligned}$$

- Subproblems: Sequential quadratic programming (SQP)

$$\begin{array}{lll} \mathbf{P_k} & \text{minimize} & \phi(x^k) + \nabla\phi(x^k)^T(x-x^k) + \frac{1}{2}[x-x^k]^\top H_k[x-x^k] \\ & \text{subject to} & f_i(x^k) + \nabla f_i(x^k)^T(x-x^k) = 0,\ i=1,\dots,s \\ & & f_i(x^k) + \nabla f_i(x^k)^T(x-x^k) = 0,\ i=s+1,\dots,m. \end{array}$$

Newton's Method Hypotheses

Let $f = h \circ c$ be PLQ convex composite, $\overline{x} \in \operatorname{dom} f$ and $\overline{y} \in \partial h(c(\overline{x}))$.

Assumptions:

(a) c is $\mathcal{C}^3$-smooth,

(b) active manifold non-degeneracy (LICQ),

(c) strict complementarity: $\ker c'(\overline{x})^T \cap \operatorname{ri} \partial h(c(\overline{x})) \neq \emptyset$,

(d) $\overline{x}$ satisfies the second-order sufficient conditions, i.e.,

$$h''(c(\overline{x}); \nabla c(\overline{x}) d) + \langle d, \nabla^2_{xx} L(\overline{x}, \overline{y}) d \rangle > 0 \quad \forall\, d \in (\ker A^\top \nabla c(\overline{x})) \backslash \{0\},$$

where the matrix A is such that the active manifold is parallel to $\ker A^T$.

Convergence of Newton's Method: B. –Engle (19)

There exists a neighborhood $\mathcal{N}$ of $(\overline{x}, \overline{y})$ such that if $(x^0, y^0) \in \mathcal{N}$, then there exists a unique sequence $\{(x^k, y^k)\}$ satisfying the optimality conditions of $\mathbf{P_k}$ with $H_k := \nabla^2_{xx} L(x^k, y^k)$ such that, for all $k \in \mathbb{N}$,

(i) $c(x^{k-1}) + \nabla c(x^{k-1})[x^k - x^{k-1}] \in$ active manifold,

(ii) $y^k \in \operatorname{ri} \partial h(c(x^{k-1}) + \nabla c(x^{k-1})[x^k - x^{k-1}])$,

(iii) $H_{k-1}[x^k - x^{k-1}] + \nabla c(x^{k-1})^\top y^k = 0$,

(iv) x^{k+1} is a strong local minimizer of $\mathbf{P_k}$.

Moreover, the sequence (x^k, y^k) converges to $(\overline{x}, \overline{y})$ at a quadratic rate.

Convergence of Newton's Method: B. –Engle (19)

There exists a neighborhood $\mathcal{N}$ of $(\overline{x}, \overline{y})$ such that if $(x^0, y^0) \in \mathcal{N}$, then there exists a unique sequence $\{(x^k, y^k)\}$ satisfying the optimality conditions of $\mathbf{P_k}$ with $H_k := \nabla^2_{xx} L(x^k, y^k)$ such that, for all $k \in \mathbb{N}$,

(i) $c(x^{k-1}) + \nabla c(x^{k-1})[x^k - x^{k-1}] \in$ active manifold,

(ii) $y^k \in \mathrm{ri}\, \partial h(c(x^{k-1}) + \nabla c(x^{k-1})[x^k - x^{k-1}])$,

(iii) $H_{k-1}[x^k - x^{k-1}] + \nabla c(x^{k-1})^\top y^k = 0$,

(iv) x^{k+1} is a strong local minimizer of $\mathbf{P_k}$.

Moreover, the sequence (x^k, y^k) converges to $(\overline{x}, \overline{y})$ at a quadratic rate.

Proof uses Robinson's *generalized equations*, Rockafellar's PLQ 2^{nd}-order theory, and Lewis' *partial smoothness* techniques.

Algorithm Globalization

$$\mathcal{P} \qquad \min_{x\in\mathbb{R}^n} f(x) := h(c(x)) + g(x),$$

where $h : \mathbb{R}^m \to \mathbb{R}$ convex, $g : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ proper, convex, str'ly cont. rel. to $\operatorname{dom} g$, and $c : \mathbb{R}^n \to \mathbb{R}^m$ is $\mathcal{C}^1$.

$$\mathcal{P}_k \qquad \min_{\|d\|\le\eta_k} h(c(x^k) + \nabla c(x^k)d) + \frac{1}{2}d^T H_k d + g(x^k + d)$$

Algorithm Globalization

$$\mathcal{P} \qquad \min_{x\in\mathbb{R}^n} f(x) := h(c(x)) + g(x),$$

where $h : \mathbb{R}^m \to \mathbb{R}$ convex, $g : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ proper, convex, str'ly cont. rel. to $\operatorname{dom} g$, and $c : \mathbb{R}^n \to \mathbb{R}^m$ is $\mathcal{C}^1$.

$$\mathcal{P}_k \qquad \min_{\|d\|\le\eta_k} h(c(x^k) + \nabla c(x^k)d) + \frac{1}{2}d^T H_k d + g(x^k + d)$$

Define

$$\Delta f(x;d) := h(c(x) + \nabla c(x)d) + g(x+d) - f(x)$$

and

$$\tilde{\Delta}_\eta f(x) := \min_{\|d\|\le\eta} \Delta f(x;d)$$

Algorithm Globalization

$$\mathcal{P} \qquad \min_{x\in\mathbb{R}^n} f(x) := h(c(x)) + g(x),$$

where $h : \mathbb{R}^m \to \mathbb{R}$ convex, $g : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ proper, convex, str'ly cont. rel. to $\operatorname{dom} g$, and $c : \mathbb{R}^n \to \mathbb{R}^m$ is $\mathcal{C}^1$.

$$\mathcal{P}_k \qquad \min_{\|d\|\le\eta_k} h(c(x^k) + \nabla c(x^k)d) + \frac{1}{2}d^T H_k d + g(x^k + d)$$

Define

$$\Delta f(x;d) := h(c(x) + \nabla c(x)d) + g(x+d) - f(x)$$

and

$$\tilde{\Delta}_\eta f(x) := \min_{\|d\|\le\eta} \Delta f(x;d)$$

Fact:

$$f'(x:d) = \lim_{t\downarrow 0} \frac{\Delta f(x;d)}{t} = \inf_{t>0} \frac{\Delta f(x;d)}{t}$$

First-Order Conditions with $\Delta f(x;d)$

TFAE

(i) $0 \in \partial f(x)$;

(ii) for all $d \in \mathbb{R}^n, \; 0 \leq f'(x;d)$;

(iii) for all $d \in \mathbb{R}^n, \; 0 \leq \Delta f(x;d)$;

(iv) $\tilde{\Delta}_1 f(x) = 0$;

(v) for all $\eta > 0, \; d = 0$ solves $\min\{\Delta f(x;d) \mid \|d\| \leq \eta\}$.

First-Order Conditions with $\Delta f(x;d)$

TFAE

(i) $0 \in \partial f(x)$;

(ii) for all $d \in \mathbb{R}^n,\ 0 \leq f'(x;d)$;

(iii) for all $d \in \mathbb{R}^n,\ 0 \leq \Delta f(x;d)$;

(iv) $\tilde{\Delta}_1 f(x) = 0$;

(v) for all $\eta > 0,\ d = 0$ solves $\min\{\Delta f(x;d) \mid \|d\| \leq \eta\}$.

Cauchy Steps

First-Order Conditions with $\Delta f(x;d)$

TFAE

(i) $0 \in \partial f(x)$;

(ii) for all $d \in \mathbb{R}^n$, $0 \leq f'(x;d)$;

(iii) for all $d \in \mathbb{R}^n$, $0 \leq \Delta f(x;d)$;

(iv) $\tilde{\Delta}_1 f(x) = 0$;

(v) for all $\eta > 0$, $d = 0$ solves $\min\{\Delta f(x;d) \mid \|d\| \leq \eta\}$.

Cauchy Steps

Sufficient Decrease Condition

For all $\epsilon > 0$ and $\eta_k > 0$, if $|\tilde{\Delta}_1 f(x)| > \epsilon$, there exists constants $\kappa_1, \kappa_2 > 0$ depending on x and ϵ such that

$$\Delta f(x^k; d^k) + \frac{1}{2} d^{k\top} H_k d^k < -\kappa_1 \min(\kappa_2, \eta_k).$$

Global Convergence (B. –Engle (19))

$$x^{k+1} := x^k + \tau_k d^k$$

- Backtracking: $\sum_{k=0}^{\infty} \frac{\Delta f(x^k; d^k)^2}{\|d^k\|_2^2} < \infty$, in particular, $\Delta f(x^k; d^k) \to 0$.

Global Convergence (B. –Engle (19))

$$x^{k+1} := x^k + \tau_k d^k$$

- Backtracking: $\displaystyle\sum_{k=0}^{\infty} \frac{\Delta f(x^k; d^k)^2}{\|d^k\|_2^2} < \infty$, in particular, $\Delta f(x^k; d^k) \to 0$.
- Weak Wolfe: $\displaystyle\sum_{k=0}^{\infty} \frac{\Delta f(x^k; d^k)^2}{\|d^k\| + \|d^k\|^2} < \infty$, in particular, $\Delta f(x^k; d^k) \to 0$.

Global Convergence (B. –Engle (19))

$$x^{k+1} := x^k + \tau_k d^k$$

- Backtracking: $\displaystyle\sum_{k=0}^{\infty} \frac{\Delta f(x^k; d^k)^2}{\|d^k\|_2^2} < \infty$, in particular, $\Delta f(x^k; d^k) \to 0$.

- Weak Wolfe: $\displaystyle\sum_{k=0}^{\infty} \frac{\Delta f(x^k; d^k)^2}{\|d^k\| + \|d^k\|^2} < \infty$, in particular, $\Delta f(x^k; d^k) \to 0$.

- Trust Region: $|\tilde{\Delta}_1 f(x^k)| \to 0$.

Complexity (Drusvyatskiy-Paquette (18))

Inexact Prox-Linear Algorithms:

- Additional Assumptions:

(i) h is L-Lipschitz: $\|h(u) - h(v)\| \leq L\|u - v\| \quad \forall\, u, v \in \mathbb{R}^m$.

(ii) c is β-Lipschitz.

Complexity (Drusvyatskiy-Paquette (18))

Inexact Prox-Linear Algorithms:

- Additional Assumptions:
  (i) h is L-Lipschitz: $\|h(u) - h(v)\| \leq L\|u - v\| \quad \forall\, u, v \in \mathbb{R}^m$.
  (ii) c is β-Lipschitz.
- Prox-Linear ingredients:

$$S_f(x) := \underset{z}{\operatorname{argmin}}\, F_t(z; x) := h(c(x) + \nabla c(x)(z - x)) + g(z) + \frac{1}{2t}\|z - x\|_2^2$$

$$\mathcal{G}_t(x) := t^{-1}\left(x - S_f(x)\right)$$

optimality $\implies \mathcal{G}_t(\overline{x}) = 0 \quad \forall\, t > 0$

Complexity (Drusvyatskiy-Paquette (18))

Inexact Prox-Linear Algorithms:

- Additional Assumptions:
  (i) h is L-Lipschitz: $\|h(u) - h(v)\| \leq L\|u - v\| \quad \forall\, u, v \in \mathbb{R}^m$.
  (ii) c is β-Lipschitz.
- Prox-Linear ingredients:

$$S_f(x) := \operatorname*{argmin}_z F_t(z; x) := h(c(x) + \nabla c(x)(z - x)) + g(z) + \frac{1}{2t}\|z - x\|_2^2$$

$$\mathcal{G}_t(x) := t^{-1}\left(x - S_f(x)\right)$$

optimality $\implies \mathcal{G}_t(\overline{x}) = 0 \quad \forall\, t > 0$

- Algorithm: $x^{k+1} \approx S_f(x^k)$ (or an ϵ_k-approx. min of $F_t(z; x^k)$)

Complexity (Drusvyatskiy-Paquette (18))

Inexact Prox-Linear Algorithms:

- Additional Assumptions:

  (i) h is L-Lipschitz: $\|h(u) - h(v)\| \leq L\|u - v\| \quad \forall\, u, v \in \mathbb{R}^m$.

  (ii) c is β-Lipschitz.

- Prox-Linear ingredients:

  $$S_f(x) := \operatorname*{argmin}_z F_t(z; x) := h(c(x) + \nabla c(x)(z - x)) + g(z) + \frac{1}{2t}\|z - x\|_2^2$$

  $$\mathcal{G}_t(x) := t^{-1}\left(x - S_f(x)\right)$$

optimality $\implies \mathcal{G}_t(\overline{x}) = 0 \quad \forall\, t > 0$

- Algorithm: $x^{k+1} \approx S_f(x^k)$ (or an ϵ_k-approx. min of $F_t(z; x^k)$)

- Convergence: If $t < (L\beta)^{-1}$, then

$$\min_{j=1,\ldots,N} \left\|\mathcal{G}_t(x^j)\right\|_2^2 \leq \frac{2(f(x^0) - \hat{f} + \sum_{j=1}^N \epsilon_j)}{tN}$$

where $\hat{f} := \liminf_k f(x^k)$.

Thank You !!

Weak Wolfe Conditions

The Weak Wolfe conditions in the convex composite case are defined for each $x \in \operatorname{dom} g$ with $\Delta f(x; d) < 0$ by choosing $0 < \sigma_1 < \sigma_2 < 1$ and $\mu > 0$ and requiring

$$f(x + td) \leq f(x) + \sigma_1 t \Delta f(x; d), \text{ and} \tag{WWI}$$

$$\sigma_2 \Delta f(x; d) \leq \frac{\Delta f(x + td; \mu d)}{\mu} \ . \tag{WWII}$$