New primal-dual subgradient methods for Convex Problems with Functional Constraints

Yurii Nesterov, CORE/INMA (UCL)

January 12, 2015 (Les Houches)

Outline

Optimization problem: simple constraints

Optimization problem: simple constraints

Consider the problem: $\min_{x \in Q} f(x),$

Optimization problem: simple constraints

Consider the problem: $\min_{x \in Q} f(x)$, where

Optimization problem: simple constraints

Consider the problem: $\min_{x \in Q} f(x)$, where

- Q is a closed convex set:

Optimization problem: simple constraints

Consider the problem: $\min_{x \in Q} f(x)$, where

- Q is a closed convex set: $x, y \in Q \Rightarrow [x, y] \subseteq Q$,

Optimization problem: simple constraints

Consider the problem: $\min_{x \in Q} f(x)$, where

- Q is a closed convex set: $x, y \in Q \Rightarrow [x, y] \subseteq Q$,
- f is a subdifferentiable on Q convex function:

Optimization problem: simple constraints

Consider the problem: $\min_{x \in Q} f(x)$, where

- Q is a closed convex set: $x, y \in Q \Rightarrow [x, y] \subseteq Q$,
- f is a subdifferentiable on Q convex function:

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle, \quad x, y \in Q, \ \nabla f(x) \in \partial f(x).$$

Optimization problem: simple constraints

Consider the problem: $\min_{x \in Q} f(x)$, where

- Q is a closed convex set: $x, y \in Q \Rightarrow [x, y] \subseteq Q$,
- f is a subdifferentiable on Q convex function:

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle, \quad x, y \in Q, \; \nabla f(x) \in \partial f(x).$$

Optimality condition:

Optimization problem: simple constraints

Consider the problem: $\min_{x \in Q} f(x)$, where

- Q is a closed convex set: $x, y \in Q \Rightarrow [x, y] \subseteq Q$,
- f is a subdifferentiable on Q convex function:

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle, \quad x, y \in Q, \; \nabla f(x) \in \partial f(x).$$

Optimality condition: point $x_* \in Q$ is optimal iff

Optimization problem: simple constraints

Consider the problem: $\min_{x \in Q} f(x)$, where

- Q is a closed convex set: $x, y \in Q \Rightarrow [x, y] \subseteq Q$,
- f is a subdifferentiable on Q convex function:

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle, \quad x, y \in Q, \ \nabla f(x) \in \partial f(x).$$

Optimality condition: point $x_* \in Q$ is optimal iff

$$\langle \nabla f(x_*), x - x_* \rangle \geq 0, \quad \forall x \in Q.$$

Optimization problem: simple constraints

Consider the problem: $\min_{x \in Q} f(x)$, where

- Q is a closed convex set: $x, y \in Q \Rightarrow [x, y] \subseteq Q$,
- f is a subdifferentiable on Q convex function:

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle, \quad x, y \in Q, \ \nabla f(x) \in \partial f(x).$$

Optimality condition: point $x_* \in Q$ is optimal iff

$$\langle \nabla f(x_*), x - x_* \rangle \geq 0, \quad \forall x \in Q.$$

Interpretation:

Optimization problem: simple constraints

Consider the problem: $\min_{x \in Q} f(x)$, where

- Q is a closed convex set: $x, y \in Q \Rightarrow [x, y] \subseteq Q$,
- f is a subdifferentiable on Q convex function:

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle, \quad x, y \in Q, \ \nabla f(x) \in \partial f(x).$$

Optimality condition: point $x_* \in Q$ is optimal iff

$$\langle \nabla f(x_*), x - x_* \rangle \geq 0, \quad \forall x \in Q.$$

Interpretation: Function increases along any feasible direction.

Optimization problem: functional constraints

Optimization problem: functional constraints

Problem:

Optimization problem: functional constraints

Problem: $\min\limits_{x \in Q} \{ f_0(x), \; f_i(x) \leq 0, \; i = 1, \dots, m \},$

Optimization problem: functional constraints

Problem: $\min\limits_{x \in Q} \{ f_0(x), \; f_i(x) \leq 0, \; i = 1, \dots, m \},$ where

Optimization problem: functional constraints

Problem: $\min\limits_{x \in Q} \{ f_0(x),\ f_i(x) \leq 0,\ i = 1, \dots, m \}$, where

- Q is a closed convex set,

Optimization problem: functional constraints

Problem: $\min\limits_{x \in Q} \{ f_0(x), \ f_i(x) \leq 0, \ i = 1, \dots, m \}$, where

- Q is a closed convex set,
- all f_i are convex and <u>subdifferentiable</u> on Q, $i = 0, \dots, m$:

Optimization problem: functional constraints

Problem: $\min\limits_{x \in Q} \{ f_0(x),\ f_i(x) \leq 0,\ i = 1, \dots, m \}$, where

- Q is a closed convex set,
- all f_i are convex and subdifferentiable on Q, $i = 0, \dots, m$:

$$f_i(y) \geq f_i(x) + \langle \nabla f_i(x), y - x \rangle, \quad x, y \in Q,\ \nabla f_i(x) \in \partial f_i(x).$$

Optimization problem: functional constraints

Problem: $\min\limits_{x \in Q} \{ f_0(x),\ f_i(x) \leq 0,\ i = 1, \dots, m \}$, where

- Q is a closed convex set,
- all f_i are convex and subdifferentiable on Q, $i = 0, \dots, m$:

$$f_i(y) \geq f_i(x) + \langle \nabla f_i(x), y - x \rangle, \quad x, y \in Q,\ \nabla f_i(x) \in \partial f_i(x).$$

Optimality condition (KKT, 1951):

Optimization problem: functional constraints

Problem: $\min\limits_{x \in Q} \{ f_0(x),\ f_i(x) \leq 0,\ i = 1, \dots, m \}$, where

- Q is a closed convex set,
- all f_i are convex and subdifferentiable on Q, $i = 0, \dots, m$:

$$f_i(y) \geq f_i(x) + \langle \nabla f_i(x), y - x \rangle, \quad x, y \in Q,\ \nabla f_i(x) \in \partial f_i(x).$$

Optimality condition (KKT, 1951): point $x_* \in Q$ is optimal iff

Optimization problem: functional constraints

Problem: $\min\limits_{x \in Q} \{ f_0(x), \; f_i(x) \leq 0, \; i = 1, \dots, m \}$, where

- Q is a closed convex set,
- all f_i are convex and subdifferentiable on Q, $i = 0, \dots, m$:

$$f_i(y) \geq f_i(x) + \langle \nabla f_i(x), y - x \rangle, \quad x, y \in Q, \; \nabla f_i(x) \in \partial f_i(x).$$

Optimality condition (KKT, 1951): point $x_* \in Q$ is optimal iff

there exist *Lagrange multipliers* $\lambda_*^{(i)} \geq 0$, $i = 1, \dots, m$, such that

Optimization problem: functional constraints

Problem: $\min\limits_{x \in Q} \{ f_0(x),\ f_i(x) \leq 0,\ i = 1, \dots, m \}$, where

- Q is a closed convex set,
- all f_i are convex and subdifferentiable on Q, $i = 0, \dots, m$:

$$f_i(y) \geq f_i(x) + \langle \nabla f_i(x), y - x \rangle, \quad x, y \in Q,\ \nabla f_i(x) \in \partial f_i(x).$$

Optimality condition (KKT, 1951): point $x_* \in Q$ is optimal iff

there exist *Lagrange multipliers* $\lambda_*^{(i)} \geq 0$, $i = 1, \dots, m$, such that

$$(1): \qquad \langle \nabla f_0(x_*) + \sum_{i=1}^{m} \lambda_*^{(i)} \nabla f_i(x_*), x - x_* \rangle \geq 0, \quad \forall x \in Q,$$

Optimization problem: functional constraints

Problem: $\min\limits_{x \in Q} \{ f_0(x),\ f_i(x) \le 0,\ i = 1, \dots, m \}$, where

- Q is a closed convex set,
- all f_i are convex and subdifferentiable on Q, $i = 0, \dots, m$:

$$f_i(y) \ge f_i(x) + \langle \nabla f_i(x), y - x \rangle, \quad x, y \in Q,\ \nabla f_i(x) \in \partial f_i(x).$$

Optimality condition (KKT, 1951): point $x_* \in Q$ is optimal iff there exist *Lagrange multipliers* $\lambda_*^{(i)} \ge 0$, $i = 1, \dots, m$, such that

$$(1): \quad \langle \nabla f_0(x_*) + \sum_{i=1}^{m} \lambda_*^{(i)} \nabla f_i(x_*), x - x_* \rangle \ge 0, \quad \forall x \in Q,$$

$$(2): \quad f_i(x_*) \le 0, \quad i = 1, \dots, m, \quad \text{(feasibility)}$$

Optimization problem: functional constraints

Problem: $\min\limits_{x \in Q} \{ f_0(x), \; f_i(x) \leq 0, \; i = 1, \dots, m \}$, where

- Q is a closed convex set,
- all f_i are convex and <u>subdifferentiable</u> on Q, $i = 0, \dots, m$:

$$f_i(y) \geq f_i(x) + \langle \nabla f_i(x), y - x \rangle, \quad x, y \in Q, \; \nabla f_i(x) \in \partial f_i(x).$$

Optimality condition (KKT, 1951): point $x_* \in Q$ is optimal iff

there exist *Lagrange multipliers* $\lambda_*^{(i)} \geq 0$, $i = 1, \dots, m$, such that

$$(1): \quad \langle \nabla f_0(x_*) + \sum_{i=1}^{m} \lambda_*^{(i)} \nabla f_i(x_*), x - x_* \rangle \geq 0, \quad \forall x \in Q,$$

$$(2): \quad f_i(x_*) \leq 0, \quad i = 1, \dots, m, \quad \text{(feasibility)}$$

$$(3): \quad \lambda_*^{(i)} f_i(x_*) = 0, \quad i = 1, \dots, m. \quad \text{(complementary slackness)}$$

Lagrange multipliers: interpretation

Lagrange multipliers: interpretation

Let $\mathcal{I} \subseteq \{1, \ldots, m\}$ be an arbitrary set of indexes.

Lagrange multipliers: interpretation

Let $\mathcal{I} \subseteq \{1, \ldots, m\}$ be an arbitrary set of indexes.

Denote $f_{\mathcal{I}}(x) = f_0(x) + \sum\limits_{i \in \mathcal{I}} \lambda_*^{(i)} f_i(x)$.

Lagrange multipliers: interpretation

Let $\mathcal{I} \subseteq \{1, \ldots, m\}$ be an arbitrary set of indexes.

Denote $f_{\mathcal{I}}(x) = f_0(x) + \sum\limits_{i \in \mathcal{I}} \lambda_*^{(i)} f_i(x)$. Consider the problem

Lagrange multipliers: interpretation

Let $\mathcal{I} \subseteq \{1, \dots, m\}$ be an arbitrary set of indexes.

Denote $f_{\mathcal{I}}(x) = f_0(x) + \sum\limits_{i \in \mathcal{I}} \lambda_*^{(i)} f_i(x)$. Consider the problem

$$\mathcal{P}_{\mathcal{I}}: \qquad \min_{x \in Q} \{ f_{\mathcal{I}}(x) : \; f_i(x) \leq 0, \; i \notin \mathcal{I} \}.$$

Lagrange multipliers: interpretation

Let $\mathcal{I} \subseteq \{1, \dots, m\}$ be an arbitrary set of indexes.

Denote $f_{\mathcal{I}}(x) = f_0(x) + \sum\limits_{i \in \mathcal{I}} \lambda_*^{(i)} f_i(x)$. Consider the problem

$$\mathcal{P}_{\mathcal{I}} : \qquad \min_{x \in Q} \{ f_{\mathcal{I}}(x) : \; f_i(x) \leq 0, \; i \notin \mathcal{I} \}.$$

Observation: in any case, x_* is the optimal solution of problem $\mathcal{P}_{\mathcal{I}}$.

Lagrange multipliers: interpretation

Let $\mathcal{I} \subseteq \{1, \dots, m\}$ be an arbitrary set of indexes.

Denote $f_{\mathcal{I}}(x) = f_0(x) + \sum\limits_{i \in \mathcal{I}} \lambda_*^{(i)} f_i(x)$. Consider the problem

$$\mathcal{P}_{\mathcal{I}} : \qquad \min_{x \in Q} \{ f_{\mathcal{I}}(x) : \; f_i(x) \leq 0, \; i \notin \mathcal{I} \}.$$

Observation: in any case, x_* is the optimal solution of problem $\mathcal{P}_{\mathcal{I}}$.

Interpretation: $\lambda_*^{(i)}$ are the *shadow prices* for resources.

Lagrange multipliers: interpretation

Let $\mathcal{I} \subseteq \{1, \ldots, m\}$ be an arbitrary set of indexes.

Denote $f_{\mathcal{I}}(x) = f_0(x) + \sum\limits_{i \in \mathcal{I}} \lambda_*^{(i)} f_i(x)$. Consider the problem

$$\mathcal{P}_{\mathcal{I}} : \qquad \min_{x \in Q} \{ f_{\mathcal{I}}(x) : \; f_i(x) \leq 0, \; i \notin \mathcal{I} \}.$$

Observation: in any case, x_* is the optimal solution of problem $\mathcal{P}_{\mathcal{I}}$.

Interpretation: $\lambda_*^{(i)}$ are the *shadow prices* for resources. **(Kantorovich, 1939)**

Lagrange multipliers: interpretation

Let $\mathcal{I} \subseteq \{1, \ldots, m\}$ be an arbitrary set of indexes.

Denote $f_{\mathcal{I}}(x) = f_0(x) + \sum\limits_{i \in \mathcal{I}} \lambda_*^{(i)} f_i(x)$. Consider the problem

$$\mathcal{P}_{\mathcal{I}} : \qquad \min_{x \in Q} \{ f_{\mathcal{I}}(x) : \; f_i(x) \leq 0, \; i \notin \mathcal{I} \}.$$

Observation: in any case, x_* is the optimal solution of problem $\mathcal{P}_{\mathcal{I}}$.

Interpretation: $\lambda_*^{(i)}$ are the *shadow prices* for resources. **(Kantorovich, 1939)**

Application examples:

Lagrange multipliers: interpretation

Let $\mathcal{I} \subseteq \{1, \ldots, m\}$ be an arbitrary set of indexes.

Denote $f_{\mathcal{I}}(x) = f_0(x) + \sum\limits_{i \in \mathcal{I}} \lambda_*^{(i)} f_i(x)$. Consider the problem

$$\mathcal{P}_{\mathcal{I}} : \qquad \min_{x \in Q} \{ f_{\mathcal{I}}(x) : \; f_i(x) \leq 0, \; i \notin \mathcal{I} \}.$$

Observation: in any case, x_* is the optimal solution of problem $\mathcal{P}_{\mathcal{I}}$.

Interpretation: $\lambda_*^{(i)}$ are the *shadow prices* for resources. **(Kantorovich, 1939)**

Application examples:

- Traffic congestion: car flows on roads $\Leftrightarrow$ size of queues.

Lagrange multipliers: interpretation

Let $\mathcal{I} \subseteq \{1, \ldots, m\}$ be an arbitrary set of indexes.

Denote $f_{\mathcal{I}}(x) = f_0(x) + \sum\limits_{i \in \mathcal{I}} \lambda_*^{(i)} f_i(x)$. Consider the problem

$$\mathcal{P}_{\mathcal{I}} : \qquad \min_{x \in Q} \{ f_{\mathcal{I}}(x) : \; f_i(x) \leq 0, \; i \notin \mathcal{I} \}.$$

Observation: in any case, x_* is the optimal solution of problem $\mathcal{P}_{\mathcal{I}}$.

Interpretation: $\lambda_*^{(i)}$ are the *shadow prices* for resources. **(Kantorovich, 1939)**

Application examples:

- Traffic congestion: car flows on roads $\Leftrightarrow$ size of queues.
- Electrical networks: currents in the wires $\Leftrightarrow$ voltage potentials, etc.

Lagrange multipliers: interpretation

Let $\mathcal{I} \subseteq \{1, \dots, m\}$ be an arbitrary set of indexes.

Denote $f_{\mathcal{I}}(x) = f_0(x) + \sum\limits_{i \in \mathcal{I}} \lambda_*^{(i)} f_i(x)$. Consider the problem

$$\mathcal{P}_{\mathcal{I}} : \qquad \min_{x \in Q} \{ f_{\mathcal{I}}(x) : \; f_i(x) \le 0, \; i \notin \mathcal{I} \}.$$

Observation: in any case, x_* is the optimal solution of problem $\mathcal{P}_{\mathcal{I}}$.

Interpretation: $\lambda_*^{(i)}$ are the *shadow prices* for resources. **(Kantorovich, 1939)**

Application examples:

- Traffic congestion: car flows on roads $\Leftrightarrow$ size of queues.
- Electrical networks: currents in the wires $\Leftrightarrow$ voltage potentials, etc.

Main question: How to compute (x_*, λ_*)?

Algebraic interpretation

Algebraic interpretation

Consider the Lagrangian $\mathcal{L}(x, \lambda) = f_0(x) + \sum_{i=1}^{m} \lambda^{(i)} f_i(x)$.

Algebraic interpretation

Consider the Lagrangian $\quad \mathcal{L}(x,\lambda) = f_0(x) + \sum\limits_{i=1}^{m} \lambda^{(i)} f_i(x)$.

Condition KKT(1): $\langle \nabla f_0(x_*) + \sum\limits_{i=1}^{m} \lambda_*^{(i)} \nabla f_i(x_*), x - x_* \rangle \geq 0$,
$\forall x \in Q$,

Algebraic interpretation

Consider the Lagrangian $\quad \mathcal{L}(x,\lambda) = f_0(x) + \sum\limits_{i=1}^{m} \lambda^{(i)} f_i(x)$.

Condition KKT(1): $\langle \nabla f_0(x_*) + \sum\limits_{i=1}^{m} \lambda_*^{(i)} \nabla f_i(x_*), x - x_* \rangle \geq 0$,

$\forall x \in Q$, implies

$$\boxed{x_* \in \operatorname{Arg}\min_{x \in Q} \mathcal{L}(x, \lambda_*).}$$

Algebraic interpretation

Consider the Lagrangian $\quad \mathcal{L}(x,\lambda) = f_0(x) + \sum\limits_{i=1}^{m} \lambda^{(i)} f_i(x).$

Condition KKT(1): $\langle \nabla f_0(x_*) + \sum\limits_{i=1}^{m} \lambda_*^{(i)} \nabla f_i(x_*), x - x_* \rangle \geq 0,$

$\forall x \in Q$, implies

$$\boxed{x_* \in \operatorname{Arg}\min_{x \in Q} \mathcal{L}(x, \lambda_*).}$$

Define the dual function $\phi(\lambda) = \min\limits_{x \in Q} \mathcal{L}(x,\lambda)$, $\lambda \geq 0$.

Algebraic interpretation

Consider the Lagrangian $\mathcal{L}(x,\lambda) = f_0(x) + \sum\limits_{i=1}^{m} \lambda^{(i)} f_i(x)$.

Condition KKT(1): $\langle \nabla f_0(x_*) + \sum\limits_{i=1}^{m} \lambda_*^{(i)} \nabla f_i(x_*), x - x_* \rangle \geq 0$,

$\forall x \in Q$, implies

$$\boxed{x_* \in \operatorname{Arg}\min_{x \in Q} \mathcal{L}(x, \lambda_*).}$$

Define the dual function $\phi(\lambda) = \min\limits_{x \in Q} \mathcal{L}(x,\lambda)$, $\lambda \geq 0$. It is concave!

Algebraic interpretation

Consider the Lagrangian $\mathcal{L}(x,\lambda) = f_0(x) + \sum\limits_{i=1}^{m} \lambda^{(i)} f_i(x)$.

Condition KKT(1): $\langle \nabla f_0(x_*) + \sum\limits_{i=1}^{m} \lambda_*^{(i)} \nabla f_i(x_*), x - x_* \rangle \geq 0$,

$\forall x \in Q$, implies

$$\boxed{x_* \in \operatorname{Arg}\min_{x \in Q} \mathcal{L}(x, \lambda_*).}$$

Define the dual function $\phi(\lambda) = \min\limits_{x \in Q} \mathcal{L}(x,\lambda)$, $\lambda \geq 0$. It is concave!

By Danskin's Theorem, $\nabla \phi(\lambda) = (f_1(x(\lambda)), \dots, f_m(x(\lambda)))$, with

$$x(\lambda) \in \operatorname{Arg}\min_{x \in Q} \mathcal{L}(x,\lambda).$$

Algebraic interpretation

Consider the Lagrangian $\mathcal{L}(x,\lambda) = f_0(x) + \sum_{i=1}^{m} \lambda^{(i)} f_i(x)$.

Condition KKT(1): $\langle \nabla f_0(x_*) + \sum_{i=1}^{m} \lambda_*^{(i)} \nabla f_i(x_*), x - x_* \rangle \geq 0$,
$\forall x \in Q$, implies

$$\boxed{x_* \in \operatorname{Arg}\min_{x \in Q} \mathcal{L}(x, \lambda_*).}$$

Define the dual function $\phi(\lambda) = \min_{x \in Q} \mathcal{L}(x,\lambda)$, $\lambda \geq 0$. It is concave!

By Danskin's Theorem, $\nabla \phi(\lambda) = (f_1(x(\lambda)), \dots, f_m(x(\lambda)))$, with

$$x(\lambda) \in \operatorname{Arg}\min_{x \in Q} \mathcal{L}(x,\lambda).$$

Conditions KKT(2,3): $f_i(x_*) \leq 0$, $\lambda_*^{(i)} f_i(x_*) = 0$, $i = 1, \dots, m$,

Algebraic interpretation

Consider the Lagrangian $\mathcal{L}(x,\lambda) = f_0(x) + \sum_{i=1}^{m} \lambda^{(i)} f_i(x)$.

Condition KKT(1): $\langle \nabla f_0(x_*) + \sum_{i=1}^{m} \lambda_*^{(i)} \nabla f_i(x_*), x - x_* \rangle \geq 0$,

$\forall x \in Q$, implies

$$\boxed{x_* \in \operatorname{Arg}\min_{x \in Q} \mathcal{L}(x, \lambda_*).}$$

Define the dual function $\phi(\lambda) = \min_{x \in Q} \mathcal{L}(x,\lambda)$, $\lambda \geq 0$. It is concave!

By Danskin's Theorem, $\nabla \phi(\lambda) = (f_1(x(\lambda)), \dots, f_m(x(\lambda)))$, with

$$x(\lambda) \in \operatorname{Arg}\min_{x \in Q} \mathcal{L}(x,\lambda).$$

Conditions KKT(2,3): $f_i(x_*) \leq 0$, $\lambda_*^{(i)} f_i(x_*) = 0$, $i = 1, \dots, m$, imply ($x_* = x(\lambda_*)$)

$$\boxed{\lambda_* \in \operatorname{Arg}\max_{\lambda \geq 0} \phi(\lambda).}$$

Algorithmic aspects

Algorithmic aspects

Main idea: solve the dual problem $\boxed{\max\limits_{\lambda \geq 0} \phi(\lambda)}$

Algorithmic aspects

Main idea: solve the dual problem $\boxed{\max_{\lambda \geq 0} \phi(\lambda)}$

by the *subgradient method*:

Algorithmic aspects

Main idea: solve the dual problem $\boxed{\max\limits_{\lambda \geq 0} \phi(\lambda)}$

by the *subgradient method*:

1. Compute $x(\lambda_k)$ and define $\nabla\phi(\lambda_k) = (f_1(x(\lambda_k)), \dots, f_m(x(\lambda_k)))$.

Algorithmic aspects

Main idea: solve the dual problem $\boxed{\max\limits_{\lambda \geq 0} \phi(\lambda)}$

by the *subgradient method*:

1. Compute $x(\lambda_k)$ and define $\nabla \phi(\lambda_k) = (f_1(x(\lambda_k)), \dots, f_m(x(\lambda_k)))$.

2. Update $\lambda_{k+1} = \text{Project}_{\mathbb{R}^n_+}(\lambda_k + h_k \nabla \phi(\lambda_k))$.

Algorithmic aspects

Main idea: solve the dual problem $\boxed{\max\limits_{\lambda \geq 0} \phi(\lambda)}$

by the *subgradient method*:

1. Compute $x(\lambda_k)$ and define $\nabla \phi(\lambda_k) = (f_1(x(\lambda_k)), \dots, f_m(x(\lambda_k)))$.

2. Update $\lambda_{k+1} = \text{Project}_{\mathbb{R}^n_+}(\lambda_k + h_k \nabla \phi(\lambda_k))$.

Stepsizes $h_k > 0$ are defined in the usual way.

Algorithmic aspects

Main idea: solve the dual problem $\boxed{\max\limits_{\lambda \geq 0} \phi(\lambda)}$

by the *subgradient method*:

1. Compute $x(\lambda_k)$ and define $\nabla \phi(\lambda_k) = (f_1(x(\lambda_k)), \dots, f_m(x(\lambda_k)))$.

2. Update $\lambda_{k+1} = \text{Project}_{\mathbb{R}^n_+}(\lambda_k + h_k \nabla \phi(\lambda_k))$.

Stepsizes $h_k > 0$ are defined in the usual way.

Main difficulties:

Algorithmic aspects

Main idea: solve the dual problem $\boxed{\max\limits_{\lambda \geq 0} \phi(\lambda)}$

by the *subgradient method*:

1. Compute $x(\lambda_k)$ and define $\nabla \phi(\lambda_k) = (f_1(x(\lambda_k)), \ldots, f_m(x(\lambda_k)))$.

2. Update $\lambda_{k+1} = \text{Project}_{\mathbb{R}^n_+}(\lambda_k + h_k \nabla \phi(\lambda_k))$.

Stepsizes $h_k > 0$ are defined in the usual way.

Main difficulties:

- Each iteration is time consuming.

Algorithmic aspects

Main idea: solve the dual problem $\boxed{\max\limits_{\lambda \geq 0} \phi(\lambda)}$

by the *subgradient method*:

1. Compute $x(\lambda_k)$ and define $\nabla\phi(\lambda_k) = (f_1(x(\lambda_k)), \dots, f_m(x(\lambda_k)))$.

2. Update $\lambda_{k+1} = \text{Project}_{\mathbb{R}^n_+}(\lambda_k + h_k \nabla\phi(\lambda_k))$.

Stepsizes $h_k > 0$ are defined in the usual way.

Main difficulties:

- Each iteration is time consuming.
- Unclear termination criterion.

Algorithmic aspects

Main idea: solve the dual problem $\boxed{\max\limits_{\lambda \geq 0} \phi(\lambda)}$

by the *subgradient method*:

1. Compute $x(\lambda_k)$ and define $\nabla \phi(\lambda_k) = (f_1(x(\lambda_k)), \dots, f_m(x(\lambda_k)))$.

2. Update $\lambda_{k+1} = \mathrm{Project}_{\mathbb{R}^n_+}(\lambda_k + h_k \nabla \phi(\lambda_k))$.

Stepsizes $h_k > 0$ are defined in the usual way.

Main difficulties:

- Each iteration is time consuming.
- Unclear termination criterion.
- Low rate of convergence

Algorithmic aspects

Main idea: solve the dual problem $\boxed{\max\limits_{\lambda \geq 0} \phi(\lambda)}$

by the *subgradient method*:

1. Compute $x(\lambda_k)$ and define $\nabla \phi(\lambda_k) = (f_1(x(\lambda_k)), \dots, f_m(x(\lambda_k)))$.

2. Update $\lambda_{k+1} = \text{Project}_{\mathbb{R}^n_+}(\lambda_k + h_k \nabla \phi(\lambda_k))$.

Stepsizes $h_k > 0$ are defined in the usual way.

Main difficulties:

- Each iteration is time consuming.
- Unclear termination criterion.
- Low rate of convergence ($O\left(\frac{1}{\epsilon^2}\right)$ upper-level iterations).

Augmented Lagrangian (1970's) [Hestenes, Powell, Rockafellar, Polyak, Bertsekas, ...]

Augmented Lagrangian (1970's) [Hestenes, Powell, Rockafellar, Polyak, Bertsekas, ...]

Define the Augmented Lagrangian

$$\widehat{\mathcal{L}}_K(x,\lambda) = f_0(x) + \frac{1}{2K}\sum_{i=1}^{m}\left(\lambda^{(i)} + Kf_i(x)\right)_+^2 - \frac{1}{2K}\|\lambda\|_2^2, \quad \lambda \in \mathbb{R}^m,$$

where $K > 0$ is a penalty parameter.

Augmented Lagrangian (1970's) [Hestenes, Powell, Rockafellar, Polyak, Bertsekas, ...]

Define the Augmented Lagrangian

$$\widehat{\mathcal{L}}_K(x,\lambda) = f_0(x) + \frac{1}{2K}\sum_{i=1}^{m}\left(\lambda^{(i)} + K f_i(x)\right)_+^2 - \frac{1}{2K}\|\lambda\|_2^2, \quad \lambda \in \mathbb{R}^m,$$

where $K > 0$ is a penalty parameter.

Consider the dual function $\hat{\phi}(\lambda) = \min\limits_{x \in Q} \widehat{\mathcal{L}}(x,\lambda)$.

Augmented Lagrangian (1970's) [Hestenes, Powell, Rockafellar, Polyak, Bertsekas, . . .]

Define the Augmented Lagrangian

$$\widehat{\mathcal{L}}_K(x,\lambda) = f_0(x) + \frac{1}{2K}\sum_{i=1}^{m}\left(\lambda^{(i)} + Kf_i(x)\right)_+^2 - \frac{1}{2K}\|\lambda\|_2^2, \quad \lambda \in \mathbb{R}^m,$$

where $K > 0$ is a penalty parameter.

Consider the dual function $\hat{\phi}(\lambda) = \min\limits_{x\in Q} \widehat{\mathcal{L}}(x,\lambda)$.

- **Main properties.** Function $\hat{\phi}$ is concave.

Augmented Lagrangian (1970's) [Hestenes, Powell, Rockafellar, Polyak, Bertsekas, ...]

Define the Augmented Lagrangian

$$\widehat{\mathcal{L}}_K(x,\lambda) = f_0(x) + \frac{1}{2K}\sum_{i=1}^{m}\left(\lambda^{(i)} + Kf_i(x)\right)_+^2 - \frac{1}{2K}\|\lambda\|_2^2, \quad \lambda \in \mathbb{R}^m,$$

where $K > 0$ is a penalty parameter.

Consider the dual function $\hat{\phi}(\lambda) = \min\limits_{x \in Q} \widehat{\mathcal{L}}(x,\lambda)$.

- **Main properties.** Function $\hat{\phi}$ is concave. Its gradient is Lipschitz continuous with constant $\frac{1}{K}$.

Augmented Lagrangian (1970's) [Hestenes, Powell, Rockafellar, Polyak, Bertsekas, ...]

Define the Augmented Lagrangian

$$\widehat{\mathcal{L}}_K(x,\lambda) = f_0(x) + \frac{1}{2K}\sum_{i=1}^{m}\left(\lambda^{(i)} + Kf_i(x)\right)_+^2 - \frac{1}{2K}\|\lambda\|_2^2, \quad \lambda \in \mathbb{R}^m,$$

where $K > 0$ is a penalty parameter.

Consider the dual function $\hat{\phi}(\lambda) = \min\limits_{x\in Q} \widehat{\mathcal{L}}(x,\lambda)$.

- **Main properties.** Function $\hat{\phi}$ is concave. Its gradient is Lipschitz continuous with constant $\frac{1}{K}$.
- Its unconstrained maximum is attained at the optimal dual solution.

Augmented Lagrangian (1970's) [Hestenes, Powell, Rockafellar, Polyak, Bertsekas, . . .]

Define the Augmented Lagrangian

$$\widehat{\mathcal{L}}_K(x,\lambda) = f_0(x) + \frac{1}{2K}\sum_{i=1}^{m}\left(\lambda^{(i)} + Kf_i(x)\right)_+^2 - \frac{1}{2K}\|\lambda\|_2^2, \quad \lambda \in \mathbb{R}^m,$$

where $K > 0$ is a penalty parameter.

Consider the dual function $\hat{\phi}(\lambda) = \min\limits_{x\in Q} \widehat{\mathcal{L}}(x,\lambda)$.

- **Main properties.** Function $\hat{\phi}$ is concave. Its gradient is Lipschitz continuous with constant $\frac{1}{K}$.
- Its unconstrained maximum is attained at the optimal dual solution.
- The corresponding point $\hat{x}(\lambda_*)$ is the optimal primal solution.

Augmented Lagrangian (1970's) [Hestenes, Powell, Rockafellar, Polyak, Bertsekas, ...]

Define the Augmented Lagrangian

$$\widehat{\mathcal{L}}_K(x,\lambda) = f_0(x) + \frac{1}{2K}\sum_{i=1}^{m}\left(\lambda^{(i)} + Kf_i(x)\right)_+^2 - \frac{1}{2K}\|\lambda\|_2^2, \quad \lambda \in \mathbb{R}^m,$$

where $K > 0$ is a penalty parameter.

Consider the dual function $\hat{\phi}(\lambda) = \min\limits_{x \in Q} \widehat{\mathcal{L}}(x,\lambda)$.

- **Main properties.** Function $\hat{\phi}$ is concave. Its gradient is Lipschitz continuous with constant $\frac{1}{K}$.
- Its unconstrained maximum is attained at the optimal dual solution.
- The corresponding point $\hat{x}(\lambda_*)$ is the optimal primal solution.

Hint: Check that the equation $\left(\lambda^{(i)} + Kf_i(x)\right)_+ = \lambda^{(i)}$ is equivalent to KKT(2,3).

Method of Augmented Lagrangians

Method of Augmented Lagrangians

Note that $\nabla \hat{\phi}(\lambda) = \frac{1}{K}\left(\lambda^{(i)} + Kf_i(x)\right)_+ - \frac{1}{K}\lambda$.

Method of Augmented Lagrangians

Note that $\nabla \hat{\phi}(\lambda) = \frac{1}{K}\left(\lambda^{(i)} + K f_i(x)\right)_+ - \frac{1}{K}\lambda$.

Therefore, the usual gradient method $\lambda_{k+1} = \lambda_k + K \nabla \hat{\phi}(\lambda_k)$ is exactly as follows:

Method of Augmented Lagrangians

Note that $\nabla \hat{\phi}(\lambda) = \frac{1}{K}\left(\lambda^{(i)} + Kf_i(x)\right)_+ - \frac{1}{K}\lambda$.

Therefore, the usual gradient method $\lambda_{k+1} = \lambda_k + K\nabla\hat{\phi}(\lambda_k)$ is exactly as follows:

Method: $\lambda_{k+1} = (\lambda_k + Kf(\hat{x}(\lambda_k)))_+$.

Method of Augmented Lagrangians

Note that $\nabla \hat{\phi}(\lambda) = \frac{1}{K} \left(\lambda^{(i)} + K f_i(x) \right)_+ - \frac{1}{K} \lambda$.

Therefore, the usual gradient method $\lambda_{k+1} = \lambda_k + K \nabla \hat{\phi}(\lambda_k)$ is exactly as follows:

Method: $\quad \lambda_{k+1} = (\lambda_k + K f(\hat{x}(\lambda_k)))_+$.

Advantage: Fast convergence of the dual process.

Method of Augmented Lagrangians

Note that $\nabla \hat{\phi}(\lambda) = \frac{1}{K}\left(\lambda^{(i)} + K f_i(x)\right)_+ - \frac{1}{K}\lambda$.

Therefore, the usual gradient method $\lambda_{k+1} = \lambda_k + K \nabla \hat{\phi}(\lambda_k)$ is exactly as follows:

Method: $\quad \lambda_{k+1} = (\lambda_k + K f(\hat{x}(\lambda_k)))_+$.

Advantage: Fast convergence of the dual process.

Disadvantages:

Method of Augmented Lagrangians

Note that $\nabla \hat{\phi}(\lambda) = \frac{1}{K} \left(\lambda^{(i)} + K f_i(x) \right)_+ - \frac{1}{K} \lambda$.

Therefore, the usual gradient method $\lambda_{k+1} = \lambda_k + K \nabla \hat{\phi}(\lambda_k)$ is exactly as follows:

Method: $\lambda_{k+1} = (\lambda_k + K f(\hat{x}(\lambda_k)))_+$.

Advantage: Fast convergence of the dual process.

Disadvantages:

- Difficult iteration.

Method of Augmented Lagrangians

Note that $\nabla \hat{\phi}(\lambda) = \frac{1}{K}\left(\lambda^{(i)} + Kf_i(x)\right)_+ - \frac{1}{K}\lambda$.

Therefore, the usual gradient method $\lambda_{k+1} = \lambda_k + K\nabla\hat{\phi}(\lambda_k)$ is exactly as follows:

Method: $\lambda_{k+1} = (\lambda_k + Kf(\hat{x}(\lambda_k)))_+$.

Advantage: Fast convergence of the dual process.

Disadvantages:

- Difficult iteration.
- Unclear termination.

Method of Augmented Lagrangians

Note that $\nabla \hat{\phi}(\lambda) = \frac{1}{K} \left(\lambda^{(i)} + K f_i(x) \right)_+ - \frac{1}{K} \lambda$.

Therefore, the usual gradient method $\lambda_{k+1} = \lambda_k + K \nabla \hat{\phi}(\lambda_k)$ is exactly as follows:

Method: $\lambda_{k+1} = (\lambda_k + K f(\hat{x}(\lambda_k)))_+$.

Advantage: Fast convergence of the dual process.

Disadvantages:

- Difficult iteration.
- Unclear termination.
- No global complexity analysis.

Method of Augmented Lagrangians

Note that $\nabla \hat{\phi}(\lambda) = \frac{1}{K} \left(\lambda^{(i)} + K f_i(x) \right)_+ - \frac{1}{K} \lambda$.

Therefore, the usual gradient method $\lambda_{k+1} = \lambda_k + K \nabla \hat{\phi}(\lambda_k)$ is exactly as follows:

Method: $\lambda_{k+1} = (\lambda_k + K f(\hat{x}(\lambda_k)))_+$.

Advantage: Fast convergence of the dual process.

Disadvantages:

- Difficult iteration.
- Unclear termination.
- No global complexity analysis.

Do we have an alternative?

Problem formulation

Problem formulation

Problem: $f^* = \inf\limits_{x \in Q} \{ f_0(x) : \; f_i(x) \leq 0, \; i = 1, \dots, m \},$

Problem formulation

Problem: $f^* = \inf\limits_{x \in Q} \{ f_0(x) : \; f_i(x) \leq 0, \; i = 1, \dots, m \}$, where

- $f_i(x)$, $i = 0, \dots, m$, are closed convex functions on Q endowed with a first-order black-box oracles,

Problem formulation

Problem: $f^* = \inf\limits_{x \in Q} \{ f_0(x) : \; f_i(x) \leq 0, \; i = 1, \dots, m \}$, where

- $f_i(x)$, $i = 0, \dots, m$, are closed convex functions on Q endowed with a first-order black-box oracles,
- $Q \subset \mathbb{E}$ is a bounded *simple* closed convex set.

Problem formulation

Problem: $f^* = \inf\limits_{x \in Q} \{ f_0(x) : \; f_i(x) \leq 0, \; i = 1, \dots, m \}$, where

- $f_i(x)$, $i = 0, \dots, m$, are closed convex functions on Q endowed with a first-order black-box oracles,
- $Q \subset \mathbb{E}$ is a bounded *simple* closed convex set. (We can solve some auxiliary optimization problems over Q.)

Problem formulation

Problem: $f^* = \inf\limits_{x \in Q} \{ f_0(x) : \; f_i(x) \leq 0, \; i = 1, \dots, m \}$, where

- $f_i(x)$, $i = 0, \dots, m$, are closed convex functions on Q endowed with a first-order black-box oracles,
- $Q \subset \mathbb{E}$ is a bounded *simple* closed convex set. (We can solve some auxiliary optimization problems over Q.)

Defining the Lagrangian

$$\mathcal{L}(x, \lambda) = f_0(x) + \sum_{i=1}^{m} \lambda^{(i)} f_i(x), \quad x \in Q, \; \lambda \in \mathbb{R}^m_+,$$

Problem formulation

Problem: $f^* = \inf\limits_{x \in Q} \{ f_0(x) : \; f_i(x) \leq 0, \; i = 1, \dots, m \}$, where

- $f_i(x)$, $i = 0, \dots, m$, are closed convex functions on Q endowed with a first-order black-box oracles,
- $Q \subset \mathbb{E}$ is a bounded *simple* closed convex set. (We can solve some auxiliary optimization problems over Q.)

Defining the Lagrangian

$$\mathcal{L}(x, \lambda) = f_0(x) + \sum_{i=1}^{m} \lambda^{(i)} f_i(x), \quad x \in Q, \; \lambda \in \mathbb{R}^m_+,$$

we can introduce the Lagrangian dual problem $\boxed{f_* \stackrel{\mathrm{def}}{=} \sup\limits_{\lambda \in \mathbb{R}^m_+} \phi(\lambda),}$

Problem formulation

Problem: $f^* = \inf\limits_{x \in Q} \{ f_0(x) : \; f_i(x) \leq 0, \; i = 1, \dots, m \}$, where

- $f_i(x)$, $i = 0, \dots, m$, are closed convex functions on Q endowed with a first-order black-box oracles,
- $Q \subset \mathbb{E}$ is a bounded *simple* closed convex set. (We can solve some auxiliary optimization problems over Q.)

Defining the Lagrangian

$$\mathcal{L}(x, \lambda) = f_0(x) + \sum_{i=1}^{m} \lambda^{(i)} f_i(x), \quad x \in Q, \; \lambda \in \mathbb{R}^m_+,$$

we can introduce the Lagrangian dual problem $\boxed{f_* \stackrel{\mathrm{def}}{=} \sup\limits_{\lambda \in \mathbb{R}^m_+} \phi(\lambda),}$

where $\phi(\lambda) \stackrel{\mathrm{def}}{=} \inf\limits_{x \in Q} \mathcal{L}(x, \lambda)$.

Problem formulation

Problem: $f^* = \inf\limits_{x \in Q} \{ f_0(x) : \ f_i(x) \leq 0, \ i = 1, \dots, m \}$, where

- $f_i(x)$, $i = 0, \dots, m$, are closed convex functions on Q endowed with a first-order black-box oracles,
- $Q \subset \mathbb{E}$ is a bounded *simple* closed convex set. (We can solve some auxiliary optimization problems over Q.)

Defining the Lagrangian

$$\mathcal{L}(x, \lambda) = f_0(x) + \sum_{i=1}^{m} \lambda^{(i)} f_i(x), \quad x \in Q, \ \lambda \in \mathbb{R}^m_+,$$

we can introduce the Lagrangian dual problem $\boxed{f_* \stackrel{\mathrm{def}}{=} \sup\limits_{\lambda \in \mathbb{R}^m_+} \phi(\lambda),}$

where $\phi(\lambda) \stackrel{\mathrm{def}}{=} \inf\limits_{x \in Q} \mathcal{L}(x, \lambda)$.

Clearly, $f^* \geq f_*$.

Problem formulation

Problem: $f^* = \inf_{x \in Q} \{ f_0(x) : \ f_i(x) \leq 0, \ i = 1, \dots, m \}$, where

- $f_i(x)$, $i = 0, \dots, m$, are closed convex functions on Q endowed with a first-order black-box oracles,
- $Q \subset \mathbb{E}$ is a bounded *simple* closed convex set. (We can solve some auxiliary optimization problems over Q.)

Defining the Lagrangian

$$\mathcal{L}(x, \lambda) = f_0(x) + \sum_{i=1}^{m} \lambda^{(i)} f_i(x), \quad x \in Q, \ \lambda \in \mathbb{R}^m_+,$$

we can introduce the Lagrangian dual problem $\boxed{f_* \overset{\text{def}}{=} \sup_{\lambda \in \mathbb{R}^m_+} \phi(\lambda),}$

where $\phi(\lambda) \overset{\text{def}}{=} \inf_{x \in Q} \mathcal{L}(x, \lambda)$.

Clearly, $f^* \geq f_*$. Later, we will show $f^* = f_*$ *algorithmically*.

Bregman distances

Bregman distances

Prox-function: $d(\cdot)$ is strongly convex on Q with parameter one:

$$d(y) \geq d(x) + \langle \nabla d(x), y - x \rangle + \frac{1}{2}\|y - x\|^2, \quad x, y \in Q.$$

Bregman distances

Prox-function: $d(\cdot)$ is strongly convex on Q with parameter one:

$$d(y) \geq d(x) + \langle \nabla d(x), y - x \rangle + \tfrac{1}{2}\|y - x\|^2, \quad x, y \in Q.$$

Denote by x_0 the prox-center of the set Q: $x_0 = \arg\min\limits_{x \in Q} d(x)$.

Bregman distances

Prox-function: $d(\cdot)$ is strongly convex on Q with parameter one:

$$d(y) \geq d(x) + \langle \nabla d(x), y - x \rangle + \frac{1}{2}\|y - x\|^2, \quad x, y \in Q.$$

Denote by x_0 the prox-center of the set Q: $x_0 = \arg\min_{x \in Q} d(x)$.

Assume $d(x_0) = 0$.

Bregman distances

Prox-function: $d(\cdot)$ is strongly convex on Q with parameter one:

$$d(y) \geq d(x) + \langle \nabla d(x), y - x \rangle + \frac{1}{2}\|y - x\|^2, \quad x, y \in Q.$$

Denote by x_0 the prox-center of the set Q: $x_0 = \arg\min\limits_{x \in Q} d(x)$.

Assume $d(x_0) = 0$.

Bregman distance:

$$\boxed{\beta(x,y) = d(y) - d(x) - \langle \nabla d(x), y - x \rangle, \; x, y \in Q.}$$

Bregman distances

Prox-function: $d(\cdot)$ is strongly convex on Q with parameter one:

$$d(y) \geq d(x) + \langle \nabla d(x), y - x \rangle + \frac{1}{2}\|y - x\|^2, \quad x, y \in Q.$$

Denote by x_0 the prox-center of the set Q: $x_0 = \arg\min_{x \in Q} d(x)$.

Assume $d(x_0) = 0$.

Bregman distance:

$$\boxed{\beta(x, y) = d(y) - d(x) - \langle \nabla d(x), y - x \rangle, \; x, y \in Q.}$$

Clearly, $\beta(x, y) \geq \frac{1}{2}\|x - y\|^2$ for all $x, y \in Q$.

Bregman distances

Prox-function: $d(\cdot)$ is strongly convex on Q with parameter one:

$$d(y) \geq d(x) + \langle \nabla d(x), y - x \rangle + \tfrac{1}{2}\|y - x\|^2, \quad x, y \in Q.$$

Denote by x_0 the prox-center of the set Q: $x_0 = \arg\min\limits_{x \in Q} d(x)$.

Assume $d(x_0) = 0$.

Bregman distance:

$$\boxed{\beta(x,y) = d(y) - d(x) - \langle \nabla d(x), y - x \rangle, \; x, y \in Q.}$$

Clearly, $\beta(x,y) \geq \frac{1}{2}\|x - y\|^2$ for all $x, y \in Q$.

Bregman mapping: for $x \in Q$, $g \in E^*$ and $h > 0$

Bregman distances

Prox-function: $d(\cdot)$ is strongly convex on Q with parameter one:

$$d(y) \geq d(x) + \langle \nabla d(x), y - x \rangle + \tfrac{1}{2}\|y - x\|^2, \quad x, y \in Q.$$

Denote by x_0 the prox-center of the set Q: $x_0 = \arg\min_{x \in Q} d(x)$.

Assume $d(x_0) = 0$.

Bregman distance:

$$\boxed{\beta(x, y) = d(y) - d(x) - \langle \nabla d(x), y - x \rangle, \; x, y \in Q.}$$

Clearly, $\beta(x, y) \geq \frac{1}{2}\|x - y\|^2$ for all $x, y \in Q$.

Bregman mapping: for $x \in Q$, $g \in E^*$ and $h > 0$ define

$$\mathcal{B}_h(x, g) = \arg\min_{y \in Q} \{ h\langle g, y - x \rangle + \beta(x, y) \}.$$

Bregman distances

Prox-function: $d(\cdot)$ is strongly convex on Q with parameter one:

$$d(y) \geq d(x) + \langle \nabla d(x), y - x \rangle + \frac{1}{2}\|y - x\|^2, \quad x, y \in Q.$$

Denote by x_0 the prox-center of the set Q: $x_0 = \arg\min\limits_{x \in Q} d(x)$.

Assume $d(x_0) = 0$.

Bregman distance:

$$\boxed{\beta(x, y) = d(y) - d(x) - \langle \nabla d(x), y - x \rangle, \; x, y \in Q.}$$

Clearly, $\beta(x, y) \geq \frac{1}{2}\|x - y\|^2$ for all $x, y \in Q$.

Bregman mapping: for $x \in Q$, $g \in E^*$ and $h > 0$ define

$$\mathcal{B}_h(x, g) = \arg\min_{y \in Q} \{ h\langle g, y - x \rangle + \beta(x, y) \}.$$

Examples: Euclidean distance, Entropy distance, etc.

Switching subgradient methods: Primal Method

Switching subgradient methods: Primal Method

Input parameter: the step size $h > 0$.

Switching subgradient methods: Primal Method

Input parameter: the step size $h > 0$.

Initialization : Compute the prox-center x_0.

Switching subgradient methods: Primal Method

Input parameter: the step size $h > 0$.

Initialization : Compute the prox-center x_0.

Iteration $k \geq 0$ **:**

Switching subgradient methods: Primal Method

Input parameter: the step size $h > 0$.

Initialization : Compute the prox-center x_0.

Iteration $k \geq 0$**:** a) Define $\mathcal{I}_k = \{i \in \{1, \dots, m\} : \; f_i(x_k) > h \|\nabla f_i(x_k)\|_*\}$.

Switching subgradient methods: Primal Method

Input parameter: the step size $h > 0$.

Initialization : Compute the prox-center x_0.

Iteration $k \geq 0$**:** a) Define $\mathcal{I}_k = \{i \in \{1,\dots,m\} : \; f_i(x_k) > h\|\nabla f_i(x_k)\|_*\}$.

b) If $\mathcal{I}_k = \emptyset$, then compute $x_{k+1} = \mathcal{B}_h\left(x_k, \frac{\nabla f_0(x_k)}{\|\nabla f_0(x_k)\|_*}\right)$.

Switching subgradient methods: Primal Method

Input parameter: the step size $h > 0$.

Initialization : Compute the prox-center x_0.

Iteration $k \geq 0$: a) Define $\mathcal{I}_k = \{i \in \{1, \dots, m\} : \ f_i(x_k) > h \|\nabla f_i(x_k)\|_* \}$.

b) If $\mathcal{I}_k = \emptyset$, then compute $x_{k+1} = \mathcal{B}_h \left(x_k, \frac{\nabla f_0(x_k)}{\|\nabla f_0(x_k)\|_*} \right)$.

c) If $\mathcal{I}_k \neq \emptyset$, then choose arbitrary $i_k \in \mathcal{I}_k$ and define

$$h_k = \frac{f_{i_k}(x_k)}{\|\nabla f_{i_k}(x_k)\|_*^2}.$$

Switching subgradient methods: Primal Method

Input parameter: the step size $h > 0$.

Initialization : Compute the prox-center x_0.

Iteration $k \geq 0$: a) Define $\mathcal{I}_k = \{i \in \{1, \dots, m\} : \; f_i(x_k) > h \|\nabla f_i(x_k)\|_*\}$.

b) If $\mathcal{I}_k = \emptyset$, then compute $x_{k+1} = \mathcal{B}_h\left(x_k, \frac{\nabla f_0(x_k)}{\|\nabla f_0(x_k)\|_*}\right)$.

c) If $\mathcal{I}_k \neq \emptyset$, then choose arbitrary $i_k \in \mathcal{I}_k$ and define

$h_k = \frac{f_{i_k}(x_k)}{\|\nabla f_{i_k}(x_k)\|_*^2}$. Compute $x_{k+1} = \mathcal{B}_{h_k}(x_k, \nabla f_{i_k}(x_k))$.

Switching subgradient methods: Primal Method

Input parameter: the step size $h > 0$.

Initialization : Compute the prox-center x_0.

Iteration $k \geq 0$: a) Define $\mathcal{I}_k = \{i \in \{1, \dots, m\} : \ f_i(x_k) > h \|\nabla f_i(x_k)\|_*\}$.

b) If $\mathcal{I}_k = \emptyset$, then compute $x_{k+1} = \mathcal{B}_h\left(x_k, \frac{\nabla f_0(x_k)}{\|\nabla f_0(x_k)\|_*}\right)$.

c) If $\mathcal{I}_k \neq \emptyset$, then choose arbitrary $i_k \in \mathcal{I}_k$ and define

$h_k = \frac{f_{i_k}(x_k)}{\|\nabla f_{i_k}(x_k)\|_*^2}$. Compute $x_{k+1} = \mathcal{B}_{h_k}(x_k, \nabla f_{i_k}(x_k))$.

After $t \geq 0$ iterations, define $\mathcal{F}_t = \{k \in \{0, \dots, t\} : \ \mathcal{I}_k = \emptyset\}$.

Switching subgradient methods: Primal Method

Input parameter: the step size $h > 0$.

Initialization : Compute the prox-center x_0.

Iteration $k \geq 0$: a) Define $\mathcal{I}_k = \{i \in \{1, \dots, m\} : \; f_i(x_k) > h \|\nabla f_i(x_k)\|_*\}$.

b) If $\mathcal{I}_k = \emptyset$, then compute $x_{k+1} = \mathcal{B}_h \left(x_k, \frac{\nabla f_0(x_k)}{\|\nabla f_0(x_k)\|_*} \right)$.

c) If $\mathcal{I}_k \neq \emptyset$, then choose arbitrary $i_k \in \mathcal{I}_k$ and define

$h_k = \frac{f_{i_k}(x_k)}{\|\nabla f_{i_k}(x_k)\|_*^2}$. Compute $x_{k+1} = \mathcal{B}_{h_k}(x_k, \nabla f_{i_k}(x_k))$.

After $t \geq 0$ iterations, define $\mathcal{F}_t = \{k \in \{0, \dots, t\} : \; \mathcal{I}_k = \emptyset\}$.

Denote $N(t) = |\mathcal{F}(t)|$.

Switching subgradient methods: Primal Method

Input parameter: the step size $h > 0$.

Initialization : Compute the prox-center x_0.

Iteration $k \geq 0$**:** a) Define $\mathcal{I}_k = \{i \in \{1, \dots, m\} : \; f_i(x_k) > h\|\nabla f_i(x_k)\|_*\}$.

b) If $\mathcal{I}_k = \emptyset$, then compute $x_{k+1} = \mathcal{B}_h\left(x_k, \frac{\nabla f_0(x_k)}{\|\nabla f_0(x_k)\|_*}\right)$.

c) If $\mathcal{I}_k \neq \emptyset$, then choose arbitrary $i_k \in \mathcal{I}_k$ and define

$h_k = \frac{f_{i_k}(x_k)}{\|\nabla f_{i_k}(x_k)\|_*^2}$. Compute $x_{k+1} = \mathcal{B}_{h_k}(x_k, \nabla f_{i_k}(x_k))$.

After $t \geq 0$ iterations, define $\mathcal{F}_t = \{k \in \{0, \dots, t\} : \; \mathcal{I}_k = \emptyset\}$.

Denote $N(t) = |\mathcal{F}(t)|$. It is possible that $N(t) = 0$.

Finding the dual multipliers

Finding the dual multipliers

if $N(t) > 0$, define the dual multipliers as follows:

Finding the dual multipliers

if $N(t) > 0$, define the dual multipliers as follows:

- $\lambda_t^{(0)} = h \sum\limits_{k \in \mathcal{F}_t} \frac{1}{\|\nabla f_0(x_k)\|_*}$,

Finding the dual multipliers

if $N(t) > 0$, define the dual multipliers as follows:

- $\lambda_t^{(0)} = h \sum\limits_{k \in \mathcal{F}_t} \frac{1}{\|\nabla f_0(x_k)\|_*}$,
- $\lambda_t^{(i)} = \frac{1}{\lambda_t^{(0)}} \sum\limits_{k \in \mathcal{A}_i(t)} h_k, \quad i = 1, \dots, m,$

Finding the dual multipliers

if $N(t) > 0$, define the dual multipliers as follows:

- $\lambda_t^{(0)} = h \sum\limits_{k \in \mathcal{F}_t} \frac{1}{\|\nabla f_0(x_k)\|_*}$,
- $\lambda_t^{(i)} = \frac{1}{\lambda_t^{(0)}} \sum\limits_{k \in \mathcal{A}_i(t)} h_k, \quad i = 1, \dots, m,$

  where $\mathcal{A}_i(t) = \{k \in \{0, \dots, t\} : \; i_k = i\}$, $0 \leq i \leq m$.

Finding the dual multipliers

if $N(t) > 0$, define the dual multipliers as follows:

- $\lambda_t^{(0)} = h \sum\limits_{k \in \mathcal{F}_t} \frac{1}{\|\nabla f_0(x_k)\|_*}$,
- $\lambda_t^{(i)} = \frac{1}{\lambda_t^{(0)}} \sum\limits_{k \in \mathcal{A}_i(t)} h_k, \quad i = 1, \dots, m,$

  where $\mathcal{A}_i(t) = \{k \in \{0, \dots, t\} : \ i_k = i\}$, $0 \le i \le m$.

Denote $S_t = \sum\limits_{k \in \mathcal{F}_t} \frac{1}{\|\nabla f_0(x_k)\|_*}$.

Finding the dual multipliers

if $N(t) > 0$, define the dual multipliers as follows:

- $\lambda_t^{(0)} = h \sum\limits_{k \in \mathcal{F}_t} \frac{1}{\|\nabla f_0(x_k)\|_*}$,
- $\lambda_t^{(i)} = \frac{1}{\lambda_t^{(0)}} \sum\limits_{k \in \mathcal{A}_i(t)} h_k, \quad i = 1, \dots, m$,

  where $\mathcal{A}_i(t) = \{k \in \{0, \dots, t\} : \ i_k = i\}$, $0 \leq i \leq m$.

Denote $S_t = \sum\limits_{k \in \mathcal{F}_t} \frac{1}{\|\nabla f_0(x_k)\|_*}$. If $\mathcal{F}_t = \emptyset$, then we define $S_t = 0$.

Finding the dual multipliers

if $N(t) > 0$, define the dual multipliers as follows:

- $\lambda_t^{(0)} = h \sum\limits_{k \in \mathcal{F}_t} \frac{1}{\|\nabla f_0(x_k)\|_*}$,
- $\lambda_t^{(i)} = \frac{1}{\lambda_t^{(0)}} \sum\limits_{k \in \mathcal{A}_i(t)} h_k, \quad i = 1, \dots, m,$

  where $\mathcal{A}_i(t) = \{k \in \{0, \dots, t\} : \; i_k = i\}$, $0 \leq i \leq m$.

Denote $S_t = \sum\limits_{k \in \mathcal{F}_t} \frac{1}{\|\nabla f_0(x_k)\|_*}$. If $\mathcal{F}_t = \emptyset$, then we define $S_t = 0$.

For proving convergence of the switching strategy,

Finding the dual multipliers

if $N(t) > 0$, define the dual multipliers as follows:

- $\lambda_t^{(0)} = h \sum\limits_{k \in \mathcal{F}_t} \frac{1}{\|\nabla f_0(x_k)\|_*}$,
- $\lambda_t^{(i)} = \frac{1}{\lambda_t^{(0)}} \sum\limits_{k \in \mathcal{A}_i(t)} h_k, \quad i = 1, \dots, m,$

  where $\mathcal{A}_i(t) = \{k \in \{0, \dots, t\} : \; i_k = i\}$, $0 \le i \le m$.

Denote $S_t = \sum\limits_{k \in \mathcal{F}_t} \frac{1}{\|\nabla f_0(x_k)\|_*}$. If $\mathcal{F}_t = \emptyset$, then we define $S_t = 0$.

For proving convergence of the switching strategy, we find an upper bound for the gap

$$\delta_t = \frac{1}{S_t} \sum_{k \in \mathcal{F}(t)} \frac{f_0(x_k)}{\|\nabla f_0(x_k)\|_*} - \phi(\lambda_t),$$

Finding the dual multipliers

if $N(t) > 0$, define the dual multipliers as follows:

- $\lambda_t^{(0)} = h \sum\limits_{k \in \mathcal{F}_t} \frac{1}{\|\nabla f_0(x_k)\|_*}$,
- $\lambda_t^{(i)} = \frac{1}{\lambda_t^{(0)}} \sum\limits_{k \in \mathcal{A}_i(t)} h_k, \quad i = 1, \dots, m$,

  where $\mathcal{A}_i(t) = \{k \in \{0, \dots, t\} : \; i_k = i\}$, $0 \le i \le m$.

Denote $S_t = \sum\limits_{k \in \mathcal{F}_t} \frac{1}{\|\nabla f_0(x_k)\|_*}$. If $\mathcal{F}_t = \emptyset$, then we define $S_t = 0$.

For proving convergence of the switching strategy, we find an upper bound for the gap

$$\delta_t = \frac{1}{S_t} \sum_{k \in \mathcal{F}(t)} \frac{f_0(x_k)}{\|\nabla f_0(x_k)\|_*} - \phi(\lambda_t),$$

assuming that $N(t) > 0$.

Convergence result

Convergence result

Main inequality:

$\lambda_t^{(0)} \delta_t$

Convergence result

Main inequality:

$$\lambda_t^{(0)} \delta_t \le r_0(x) + \frac{1}{2} N(t) h^2 - \frac{1}{2}(t - N(t)) h^2$$

Convergence result

Main inequality:

$$\lambda_t^{(0)} \delta_t \leq r_0(x) + \tfrac{1}{2} N(t) h^2 - \tfrac{1}{2}(t - N(t)) h^2 = r_0(x) - \tfrac{1}{2} t h^2 + N(t) h^2.$$

Convergence result

Main inequality:

$$\lambda_t^{(0)}\delta_t \leq r_0(x) + \tfrac{1}{2}N(t)h^2 - \tfrac{1}{2}(t - N(t))h^2 = r_0(x) - \tfrac{1}{2}th^2 + N(t)h^2.$$

Denote $D = \max\limits_{x \in Q} r_0(x)$.

Convergence result

Main inequality:

$$\lambda_t^{(0)}\delta_t \le r_0(x) + \tfrac{1}{2}N(t)h^2 - \tfrac{1}{2}(t - N(t))h^2 = r_0(x) - \tfrac{1}{2}th^2 + N(t)h^2.$$

Denote $D = \max\limits_{x \in Q} r_0(x)$.

Theorem. If the number $t \ge \frac{2}{h^2}D$,

Convergence result

Main inequality:

$$\lambda_t^{(0)}\delta_t \le r_0(x) + \tfrac{1}{2}N(t)h^2 - \tfrac{1}{2}(t - N(t))h^2 = r_0(x) - \tfrac{1}{2}th^2 + N(t)h^2.$$

Denote $D = \max\limits_{x \in Q} r_0(x)$.

Theorem. If the number $t \ge \frac{2}{h^2}D$, then $\mathcal{F}(t) \ne \emptyset$.

Convergence result

Main inequality:

$$\lambda_t^{(0)}\delta_t \leq r_0(x) + \frac{1}{2}N(t)h^2 - \frac{1}{2}(t - N(t))h^2 = r_0(x) - \frac{1}{2}th^2 + N(t)h^2.$$

Denote $D = \max\limits_{x \in Q} r_0(x)$.

Theorem. If the number $t \geq \frac{2}{h^2}D$, then $\mathcal{F}(t) \neq \emptyset$.

In this case $\quad \delta_t \leq Mh$

Convergence result

Main inequality:

$$\lambda_t^{(0)}\delta_t \leq r_0(x) + \frac{1}{2}N(t)h^2 - \frac{1}{2}(t - N(t))h^2 = r_0(x) - \frac{1}{2}th^2 + N(t)h^2.$$

Denote $D = \max\limits_{x \in Q} r_0(x)$.

Theorem. If the number $t \geq \frac{2}{h^2}D$, then $\mathcal{F}(t) \neq \emptyset$.

In this case $\quad \delta_t \leq Mh \quad$ and $\quad \max\limits_{1 \leq i \leq m} f_i(x_k) \leq Mh, \; k \in \mathcal{F}(t)$

where $M = \max\limits_{0 \leq k \leq t} \max\limits_{0 \leq i \leq m} \|\nabla f_i(x_k)\|_*$.

Convergence result

Main inequality:

$$\lambda_t^{(0)}\delta_t \leq r_0(x) + \frac{1}{2}N(t)h^2 - \frac{1}{2}(t - N(t))h^2 = r_0(x) - \frac{1}{2}th^2 + N(t)h^2.$$

Denote $D = \max\limits_{x \in Q} r_0(x)$.

Theorem. If the number $t \geq \frac{2}{h^2}D$, then $\mathcal{F}(t) \neq \emptyset$.

In this case $\quad \delta_t \leq Mh \quad$ and $\quad \max\limits_{1 \leq i \leq m} f_i(x_k) \leq Mh$, $k \in \mathcal{F}(t)$

where $M = \max\limits_{0 \leq k \leq t} \max\limits_{0 \leq i \leq m} \|\nabla f_i(x_k)\|_*$.

Proof:

Convergence result

Main inequality:

$$\lambda_t^{(0)} \delta_t \le r_0(x) + \tfrac{1}{2} N(t) h^2 - \tfrac{1}{2}(t - N(t)) h^2 = r_0(x) - \tfrac{1}{2} t h^2 + N(t) h^2.$$

Denote $D = \max\limits_{x \in Q} r_0(x)$.

Theorem. If the number $t \ge \frac{2}{h^2} D$, then $\mathcal{F}(t) \neq \emptyset$.

In this case $\quad \delta_t \le Mh \quad$ and $\quad \max\limits_{1 \le i \le m} f_i(x_k) \le Mh, \; k \in \mathcal{F}(t)$

where $M = \max\limits_{0 \le k \le t} \max\limits_{0 \le i \le m} \|\nabla f_i(x_k)\|_*$.

Proof: If $\mathcal{F}(t) = \emptyset$, then $N(t) = 0$.

Convergence result

Main inequality:

$$\lambda_t^{(0)}\delta_t \leq r_0(x) + \frac{1}{2}N(t)h^2 - \frac{1}{2}(t - N(t))h^2 = r_0(x) - \frac{1}{2}th^2 + N(t)h^2.$$

Denote $D = \max\limits_{x \in Q} r_0(x)$.

Theorem. If the number $t \geq \frac{2}{h^2}D$, then $\mathcal{F}(t) \neq \emptyset$.

In this case $\quad \delta_t \leq Mh \quad$ and $\quad \max\limits_{1 \leq i \leq m} f_i(x_k) \leq Mh, \; k \in \mathcal{F}(t)$

where $M = \max\limits_{0 \leq k \leq t} \max\limits_{0 \leq i \leq m} \|\nabla f_i(x_k)\|_*$.

Proof: If $\mathcal{F}(t) = \emptyset$, then $N(t) = 0$. Consequently, $\lambda_t^{(0)} = 0$.

Convergence result

Main inequality:

$$\lambda_t^{(0)}\delta_t \leq r_0(x) + \tfrac{1}{2}N(t)h^2 - \tfrac{1}{2}(t - N(t))h^2 = r_0(x) - \tfrac{1}{2}th^2 + N(t)h^2.$$

Denote $D = \max\limits_{x \in Q} r_0(x)$.

Theorem. If the number $t \geq \frac{2}{h^2}D$, then $\mathcal{F}(t) \neq \emptyset$.

In this case $\quad \delta_t \leq Mh \quad$ and $\quad \max\limits_{1 \leq i \leq m} f_i(x_k) \leq Mh, \; k \in \mathcal{F}(t)$

where $M = \max\limits_{0 \leq k \leq t} \max\limits_{0 \leq i \leq m} \|\nabla f_i(x_k)\|_*$.

Proof: If $\mathcal{F}(t) = \emptyset$, then $N(t) = 0$. Consequently, $\lambda_t^{(0)} = 0$. This is impossible for t big enough.

Convergence result

Main inequality:

$$\lambda_t^{(0)}\delta_t \leq r_0(x) + \tfrac{1}{2}N(t)h^2 - \tfrac{1}{2}(t - N(t))h^2 = r_0(x) - \tfrac{1}{2}th^2 + N(t)h^2.$$

Denote $D = \max\limits_{x \in Q} r_0(x)$.

Theorem. If the number $t \geq \frac{2}{h^2}D$, then $\mathcal{F}(t) \neq \emptyset$.

In this case $\quad \delta_t \leq Mh \quad$ and $\quad \max\limits_{1 \leq i \leq m} f_i(x_k) \leq Mh$, $k \in \mathcal{F}(t)$

where $M = \max\limits_{0 \leq k \leq t} \max\limits_{0 \leq i \leq m} \|\nabla f_i(x_k)\|_*$.

Proof: If $\mathcal{F}(t) = \emptyset$, then $N(t) = 0$. Consequently, $\lambda_t^{(0)} = 0$. This is impossible for t big enough.

Finally, $\lambda_t^{(0)} \geq \frac{h}{M}N(t)$.

Convergence result

Main inequality:

$$\lambda_t^{(0)}\delta_t \leq r_0(x) + \tfrac{1}{2}N(t)h^2 - \tfrac{1}{2}(t - N(t))h^2 = r_0(x) - \tfrac{1}{2}th^2 + N(t)h^2.$$

Denote $D = \max\limits_{x \in Q} r_0(x)$.

Theorem. If the number $t \geq \frac{2}{h^2}D$, then $\mathcal{F}(t) \neq \emptyset$.

In this case $\quad \delta_t \leq Mh \quad$ and $\quad \max\limits_{1 \leq i \leq m} f_i(x_k) \leq Mh$, $k \in \mathcal{F}(t)$

where $M = \max\limits_{0 \leq k \leq t} \max\limits_{0 \leq i \leq m} \|\nabla f_i(x_k)\|_*$.

Proof: If $\mathcal{F}(t) = \emptyset$, then $N(t) = 0$. Consequently, $\lambda_t^{(0)} = 0$. This is impossible for t big enough.

Finally, $\lambda_t^{(0)} \geq \frac{h}{M}N(t)$. Therefore, if t is big enough,

Convergence result

Main inequality:

$$\lambda_t^{(0)}\delta_t \le r_0(x) + \frac{1}{2}N(t)h^2 - \frac{1}{2}(t - N(t))h^2 = r_0(x) - \frac{1}{2}th^2 + N(t)h^2.$$

Denote $D = \max\limits_{x \in Q} r_0(x)$.

Theorem. If the number $t \ge \frac{2}{h^2}D$, then $\mathcal{F}(t) \neq \emptyset$.

In this case $\quad \delta_t \le Mh \quad$ and $\quad \max\limits_{1\le i\le m} f_i(x_k) \le Mh$, $k \in \mathcal{F}(t)$

where $M = \max\limits_{0\le k\le t}\max\limits_{0\le i\le m} \|\nabla f_i(x_k)\|_*$.

Proof: If $\mathcal{F}(t) = \emptyset$, then $N(t) = 0$. Consequently, $\lambda_t^{(0)} = 0$. This is impossible for t big enough.

Finally, $\lambda_t^{(0)} \ge \frac{h}{M}N(t)$. Therefore, if t is big enough, then

$$\delta_t \le \frac{N(t)h^2}{\lambda_t^{(0)}}$$

Convergence result

Main inequality:

$$\lambda_t^{(0)}\delta_t \leq r_0(x) + \tfrac{1}{2}N(t)h^2 - \tfrac{1}{2}(t - N(t))h^2 = r_0(x) - \tfrac{1}{2}th^2 + N(t)h^2.$$

Denote $D = \max\limits_{x \in Q} r_0(x)$.

Theorem. If the number $t \geq \frac{2}{h^2}D$, then $\mathcal{F}(t) \neq \emptyset$.

In this case $\quad \delta_t \leq Mh \quad$ and $\quad \max\limits_{1 \leq i \leq m} f_i(x_k) \leq Mh$, $k \in \mathcal{F}(t)$

where $M = \max\limits_{0 \leq k \leq t} \max\limits_{0 \leq i \leq m} \| \nabla f_i(x_k) \|_*$.

Proof: If $\mathcal{F}(t) = \emptyset$, then $N(t) = 0$. Consequently, $\lambda_t^{(0)} = 0$. This is impossible for t big enough.

Finally, $\lambda_t^{(0)} \geq \frac{h}{M}N(t)$. Therefore, if t is big enough, then

$$\delta_t \leq \frac{N(t)h^2}{\lambda_t^{(0)}} \leq Mh. \quad \square$$

Dual subgradient method

Dual subgradient method

Define the *averaging coefficients* $\{a_k\}_{k \geq 0}$, and nondecreasing *scaling coefficients* $\{\beta_k\}_{k \geq 0}$. Denote $A_t = \sum\limits_{k=0}^{t} a_k$.

Dual subgradient method

Define the *averaging coefficients* $\{a_k\}_{k\geq 0}$, and nondecreasing *scaling coefficients* $\{\beta_k\}_{k\geq 0}$. Denote $A_t = \sum\limits_{k=0}^{t} a_k$.

Initialization : Define $\ell_0(x) \equiv 0$, $x \in Q$.

Dual subgradient method

Define the *averaging coefficients* $\{a_k\}_{k \geq 0}$, and nondecreasing *scaling coefficients* $\{\beta_k\}_{k \geq 0}$. Denote $A_t = \sum\limits_{k=0}^{t} a_k$.

Initialization : Define $\ell_0(x) \equiv 0$, $x \in Q$.

Iteration $k \geq 0$ **:** a) Compute $x_k = \arg\min\limits_{x \in Q} \{\ell_k(x) + \beta_k d(x)\}$.

Dual subgradient method

Define the *averaging coefficients* $\{a_k\}_{k \geq 0}$, and nondecreasing *scaling coefficients* $\{\beta_k\}_{k \geq 0}$. Denote $A_t = \sum\limits_{k=0}^{t} a_k$.

Initialization : Define $\ell_0(x) \equiv 0$, $x \in Q$.

Iteration $k \geq 0$**:** a) Compute $x_k = \arg\min\limits_{x \in Q} \{\ell_k(x) + \beta_k d(x)\}$.

b) Define $\mathcal{I}_k = \{i \in [1:m] : \; f^{(i)}(x_k) \geq \epsilon\}$.

Dual subgradient method

Define the *averaging coefficients* $\{a_k\}_{k\geq 0}$, and nondecreasing *scaling coefficients* $\{\beta_k\}_{k\geq 0}$. Denote $A_t = \sum\limits_{k=0}^{t} a_k$.

Initialization : Define $\ell_0(x) \equiv 0$, $x \in Q$.

Iteration $k \geq 0$**:** a) Compute $x_k = \arg\min\limits_{x\in Q} \{\ell_k(x) + \beta_k d(x)\}$.

b) Define $\mathcal{I}_k = \{i \in [1:m] : \; f^{(i)}(x_k) \geq \epsilon\}$.

c) If $\mathcal{I}_k = \emptyset$, then $\ell_{k+1}(x) = \ell_k(x) + a_k[f^{(0)}(x_k) + \langle \nabla f^{(0)}(x_k), x - x_k\rangle]$.

Dual subgradient method

Define the *averaging coefficients* $\{a_k\}_{k\geq 0}$, and nondecreasing *scaling coefficients* $\{\beta_k\}_{k\geq 0}$. Denote $A_t = \sum\limits_{k=0}^{t} a_k$.

Initialization : Define $\ell_0(x) \equiv 0$, $x \in Q$.

Iteration $k \geq 0$**:** a) Compute $x_k = \arg\min\limits_{x\in Q} \{\ell_k(x) + \beta_k d(x)\}$.

b) Define $\mathcal{I}_k = \{i \in [1:m] : \; f^{(i)}(x_k) \geq \epsilon\}$.

c) If $\mathcal{I}_k = \emptyset$, then $\ell_{k+1}(x) = \ell_k(x) + a_k[f^{(0)}(x_k) + \langle \nabla f^{(0)}(x_k), x - x_k\rangle]$.

d) If $\mathcal{I}_k \neq \emptyset$, then choose arbitrary $i_k \in \mathcal{I}_k$ and define

$$\ell_{k+1}(x) = \ell_k(x) + a_k[f^{(i_k)}(x_k) + \langle \nabla f^{(i_k)}(x_k), x - x_k\rangle].$$

Convergence result

Convergence result

Define $\mathcal{A}_0(t) = \{k \in [0:t] : \mathcal{I}_k = \emptyset\}$, $N(t) = |\mathcal{A}_0(t)|$, and

$$\sigma_t = \sum_{k \in \mathcal{A}_0(t)} a_k, \quad \lambda_t^{(i)} = \frac{1}{\sigma_t} \sum_{k \in \mathcal{A}_i(t)} a_k, \quad i = 1, \dots, m,$$

where $\mathcal{A}_i(t) = \{k \in [0:t] : i_k = i\}$, $1 \leq i \leq m$.

Convergence result

Define $\mathcal{A}_0(t) = \{k \in [0:t]: \mathcal{I}_k = \emptyset\}$, $N(t) = |\mathcal{A}_0(t)|$, and

$$\sigma_t = \sum_{k \in \mathcal{A}_0(t)} a_k, \quad \lambda_t^{(i)} = \frac{1}{\sigma_t} \sum_{k \in \mathcal{A}_i(t)} a_k, \quad i = 1, \dots, m,$$

where $\mathcal{A}_i(t) = \{k \in [0:t]: i_k = i\}$, $1 \le i \le m$.

If $N(t) > 0$, then define the gap $\delta_t = \frac{1}{\sigma_t} \sum\limits_{k \in \mathcal{F}(t)} a_k f^{(0)}(x_k) - \phi(\lambda_t)$.

Convergence result

Define $\mathcal{A}_0(t) = \{k \in [0:t] : \mathcal{I}_k = \emptyset\}$, $N(t) = |\mathcal{A}_0(t)|$, and

$$\sigma_t = \sum_{k \in \mathcal{A}_0(t)} a_k, \quad \lambda_t^{(i)} = \frac{1}{\sigma_t} \sum_{k \in \mathcal{A}_i(t)} a_k, \quad i = 1, \dots, m,$$

where $\mathcal{A}_i(t) = \{k \in [0:t] : i_k = i\}$, $1 \le i \le m$.

If $N(t) > 0$, then define the gap $\delta_t = \frac{1}{\sigma_t} \sum\limits_{k \in \mathcal{F}(t)} a_k f^{(0)}(x_k) - \phi(\lambda_t)$.

Denote $D = \max\limits_{x \in Q} d(x)$.

Convergence result

Define $\mathcal{A}_0(t) = \{k \in [0:t] : \mathcal{I}_k = \emptyset\}$, $N(t) = |\mathcal{A}_0(t)|$, and

$$\sigma_t = \sum_{k \in \mathcal{A}_0(t)} a_k, \quad \lambda_t^{(i)} = \frac{1}{\sigma_t} \sum_{k \in \mathcal{A}_i(t)} a_k, \quad i = 1, \ldots, m,$$

where $\mathcal{A}_i(t) = \{k \in [0:t] : i_k = i\}$, $1 \le i \le m$.

If $N(t) > 0$, then define the gap $\delta_t = \frac{1}{\sigma_t} \sum\limits_{k \in \mathcal{F}(t)} a_k f^{(0)}(x_k) - \phi(\lambda_t)$.

Denote $D = \max\limits_{x \in Q} d(x)$.

T. Let all subgradients be bounded by M. Then for any $t \ge 0$

$$\sigma_t \cdot (\delta_t - \epsilon) + A_t \epsilon \le \beta_{t+1} D + \frac{1}{2} M^2 \sum_{k=0}^{t} \frac{a_k^2}{\beta_k}.$$

Convergence result

Define $\mathcal{A}_0(t) = \{k \in [0:t] : \mathcal{I}_k = \emptyset\}$, $N(t) = |\mathcal{A}_0(t)|$, and

$$\sigma_t = \sum_{k \in \mathcal{A}_0(t)} a_k, \quad \lambda_t^{(i)} = \frac{1}{\sigma_t} \sum_{k \in \mathcal{A}_i(t)} a_k, \quad i = 1, \dots, m,$$

where $\mathcal{A}_i(t) = \{k \in [0:t] : i_k = i\}$, $1 \leq i \leq m$.

If $N(t) > 0$, then define the gap $\delta_t = \frac{1}{\sigma_t} \sum\limits_{k \in \mathcal{F}(t)} a_k f^{(0)}(x_k) - \phi(\lambda_t)$.

Denote $D = \max\limits_{x \in Q} d(x)$.

T. Let all subgradients be bounded by M. Then for any $t \geq 0$

$$\sigma_t \cdot (\delta_t - \epsilon) + A_t \epsilon \leq \beta_{t+1} D + \frac{1}{2} M^2 \sum_{k=0}^{t} \frac{a_k^2}{\beta_k}.$$

If $A_t \epsilon > \beta_{t+1} D + \frac{1}{2} M^2 \sum\limits_{k=0}^{t} \frac{a_k^2}{\beta_k}$, then $\lambda_t^{(0)} > 0$ and $\delta_t \leq \epsilon$.

Convergence result

Define $\mathcal{A}_0(t) = \{k \in [0:t] : \ \mathcal{I}_k = \emptyset\}$, $N(t) = |\mathcal{A}_0(t)|$, and

$$\sigma_t = \sum_{k \in \mathcal{A}_0(t)} a_k, \quad \lambda_t^{(i)} = \frac{1}{\sigma_t} \sum_{k \in \mathcal{A}_i(t)} a_k, \quad i = 1, \dots, m,$$

where $\mathcal{A}_i(t) = \{k \in [0:t] : \ i_k = i\}$, $1 \le i \le m$.

If $N(t) > 0$, then define the gap $\delta_t = \frac{1}{\sigma_t} \sum\limits_{k \in \mathcal{F}(t)} a_k f^{(0)}(x_k) - \phi(\lambda_t)$.

Denote $D = \max\limits_{x \in Q} d(x)$.

T. Let all subgradients be bounded by M. Then for any $t \ge 0$

$$\sigma_t \cdot (\delta_t - \epsilon) + A_t \epsilon \le \beta_{t+1} D + \frac{1}{2} M^2 \sum_{k=0}^{t} \frac{a_k^2}{\beta_k}.$$

If $A_t \epsilon > \beta_{t+1} D + \frac{1}{2} M^2 \sum\limits_{k=0}^{t} \frac{a_k^2}{\beta_k}$, then $\lambda_t^{(0)} > 0$ and $\delta_t \le \epsilon$.

Example: $a_t \equiv 1$, $\beta_t \approx \sqrt{t} \quad \Rightarrow \quad t \approx O\left(\frac{1}{\epsilon^2}\right)$.

Quasi-monotone method

Quasi-monotone method

Initialization : Define $\ell_0(x) \equiv 0$, $x_0 = \bowtie$, and $\sigma_0 = 0$.
Iteration $t \geq 0$ **:**

Quasi-monotone method

Initialization : Define $\ell_0(x) \equiv 0$, $x_0 = \bowtie$, and $\sigma_0 = 0$.

Iteration $t \geq 0$ **:**

a) Set $v_t = \arg\min\limits_{x \in Q} \{\ell_t(x) + \beta_t d(x)\}$, $\mathcal{I}_t \stackrel{\mathrm{def}}{=} \{i \in [1:m] : \; f^{(i)}(v_t) \geq \epsilon\}$.

Quasi-monotone method

Initialization : Define $\ell_0(x) \equiv 0$, $x_0 = \bowtie$, and $\sigma_0 = 0$.

Iteration $t \geq 0$ **:**

a) Set $v_t = \arg\min\limits_{x \in Q} \{\ell_t(x) + \beta_t d(x)\}$, $\mathcal{I}_t \stackrel{\mathrm{def}}{=} \{i \in [1:m] : \ f^{(i)}(v_t) \geq \epsilon\}$.

b) If $\mathcal{I}_t \neq \emptyset$, then set $x_{t+1} = x_t$, $\sigma_{t+1} = \sigma_t$, and choose arbitrary $i_t \in \mathcal{I}_t$.

Update $\ell_{t+1}(x) = \ell_t(x) + a_{t+1}[f^{(i_t)}(v_t) + \langle \nabla f^{(i_t)}(v_t), x - v_t \rangle]$.

Quasi-monotone method

Initialization : Define $\ell_0(x) \equiv 0$, $x_0 = \bowtie$, and $\sigma_0 = 0$.

Iteration $t \geq 0$ **:**

a) Set $v_t = \arg\min_{x \in Q} \{\ell_t(x) + \beta_t d(x)\}$, $\mathcal{I}_t \stackrel{\mathrm{def}}{=} \{i \in [1:m] : \; f^{(i)}(v_t) \geq \epsilon\}$.

b) If $\mathcal{I}_t \neq \emptyset$, then set $x_{t+1} = x_t$, $\sigma_{t+1} = \sigma_t$, and choose arbitrary $i_t \in \mathcal{I}_t$.

Update $\ell_{t+1}(x) = \ell_t(x) + a_{t+1}[f^{(i_t)}(v_t) + \langle \nabla f^{(i_t)}(v_t), x - v_t \rangle]$.

c) Otherwise, $\sigma_{t+1} = \sigma_t + a_{t+1}$, $\tau_t = \frac{a_{t+1}}{\sigma_{t+1}}$, $x_{t+1} = (1 - \tau_t)x_t + \tau_t v_t$.

Update $\ell_{t+1}(x) = \ell_t(x) + a_{t+1}[f^{(0)}(x_{t+1}) + \langle \nabla f^{(0)}(x_{t+1}), x - x_{t+1} \rangle]$.

Quasi-monotone method

Initialization : Define $\ell_0(x) \equiv 0$, $x_0 = \bowtie$, and $\sigma_0 = 0$.

Iteration $t \geq 0$:

a) Set $v_t = \arg\min\limits_{x \in Q} \{\ell_t(x) + \beta_t d(x)\}$, $\mathcal{I}_t \stackrel{\mathrm{def}}{=} \{i \in [1:m]: \; f^{(i)}(v_t) \geq \epsilon\}$.

b) If $\mathcal{I}_t \neq \emptyset$, then set $x_{t+1} = x_t$, $\sigma_{t+1} = \sigma_t$, and choose arbitrary $i_t \in \mathcal{I}_t$.

Update $\ell_{t+1}(x) = \ell_t(x) + a_{t+1}[f^{(i_t)}(v_t) + \langle \nabla f^{(i_t)}(v_t), x - v_t \rangle]$.

c) Otherwise, $\sigma_{t+1} = \sigma_t + a_{t+1}$, $\tau_t = \frac{a_{t+1}}{\sigma_{t+1}}$, $x_{t+1} = (1-\tau_t)x_t + \tau_t v_t$.

Update $\ell_{t+1}(x) = \ell_t(x) + a_{t+1}[f^{(0)}(x_{t+1}) + \langle \nabla f^{(0)}(x_{t+1}), x - x_{t+1} \rangle]$.

Operation $x_0 = \bowtie \in \mathbb{E}$ indicates that x_0 is not chosen yet.

Convergence result

Convergence result

Theorem. 1. All points $x_t \neq x_{\bowtie}$ are ϵ-feasible.

Convergence result

Theorem. 1. All points $x_t \neq x_{\bowtie}$ are ϵ-feasible.

2. If all subgradients are bounded by M, then $\forall x \in Q$, $t \geq 0$,

$$\sigma_t(f^{(0)}(x_t) - \epsilon) + A_t \epsilon \quad \leq \quad \ell_t(x) + \beta_t d(x) + \tfrac{1}{2} M^2 \sum_{k=1}^{t} \frac{a_k^2}{\beta_{k-1}}.$$

Convergence result

Theorem. 1. All points $x_t \neq x_{\bowtie}$ are ϵ-feasible.

2. If all subgradients are bounded by M, then $\forall x \in Q,\ t \geq 0$,

$$\sigma_t(f^{(0)}(x_t) - \epsilon) + A_t\epsilon \quad \leq \quad \ell_t(x) + \beta_t d(x) + \tfrac{1}{2}M^2 \sum_{k=1}^{t} \frac{a_k^2}{\beta_{k-1}}.$$

3. No later than $A_t\epsilon > \beta_t D + \frac{1}{2}M^2 \sum\limits_{k=1}^{t} \frac{a_k^2}{\beta_{k-1}}$, we get $\sigma_t > 0$ and

Convergence result

Theorem. 1. All points $x_t \neq x_{\bowtie}$ are ϵ-feasible.

2. If all subgradients are bounded by M, then $\forall x \in Q,\ t \geq 0$,

$$\sigma_t(f^{(0)}(x_t) - \epsilon) + A_t\epsilon \quad \leq \quad \ell_t(x) + \beta_t d(x) + \tfrac{1}{2} M^2 \sum_{k=1}^{t} \frac{a_k^2}{\beta_{k-1}}.$$

3. No later than $A_t\epsilon > \beta_t D + \frac{1}{2} M^2 \sum\limits_{k=1}^{t} \frac{a_k^2}{\beta_{k-1}}$, we get $\sigma_t > 0$ and

$$f(x_t) - \phi(\lambda_t) \quad \leq \quad f(x_t) - \frac{1}{\sigma_t} \min_{x \in Q} \ell_t(x) \ \leq \ \epsilon.$$

Convergence result

Theorem. 1. All points $x_t \neq x_{\bowtie}$ are ϵ-feasible.

2. If all subgradients are bounded by M, then $\forall x \in Q$, $t \geq 0$,

$$\sigma_t(f^{(0)}(x_t) - \epsilon) + A_t \epsilon \quad \leq \quad \ell_t(x) + \beta_t d(x) + \tfrac{1}{2} M^2 \sum_{k=1}^{t} \frac{a_k^2}{\beta_{k-1}}.$$

3. No later than $A_t \epsilon > \beta_t D + \frac{1}{2} M^2 \sum\limits_{k=1}^{t} \frac{a_k^2}{\beta_{k-1}}$, we get $\sigma_t > 0$ and

$$f(x_t) - \phi(\lambda_t) \quad \leq \quad f(x_t) - \frac{1}{\sigma_t} \min_{x \in Q} \; \ell_t(x) \; \leq \; \epsilon.$$

Example: $a_t \equiv 1$, $\beta_t \approx \sqrt{t} \quad \Rightarrow \quad t \approx O\left(\frac{1}{\epsilon^2}\right)$.

Convergence result

Theorem. 1. All points $x_t \neq x_{\bowtie}$ are ϵ-feasible.

2. If all subgradients are bounded by M, then $\forall x \in Q,\ t \geq 0$,

$$\sigma_t(f^{(0)}(x_t) - \epsilon) + A_t \epsilon \quad \leq \quad \ell_t(x) + \beta_t d(x) + \tfrac{1}{2} M^2 \sum_{k=1}^{t} \frac{a_k^2}{\beta_{k-1}}.$$

3. No later than $A_t \epsilon > \beta_t D + \frac{1}{2} M^2 \sum\limits_{k=1}^{t} \frac{a_k^2}{\beta_{k-1}}$, we get $\sigma_t > 0$ and

$$f(x_t) - \phi(\lambda_t) \quad \leq \quad f(x_t) - \frac{1}{\sigma_t} \min_{x \in Q} \ \ell_t(x) \ \leq \ \epsilon.$$

Example: $a_t \equiv 1,\ \beta_t \approx \sqrt{t} \quad \Rightarrow \quad t \approx O\left(\frac{1}{\epsilon^2}\right)$.

NB: this is true for the whole sequence!

Conclusion

Conclusion

1. Optimal primal-dual solution can be approximated by simple switching subgradient schemes.

Conclusion

1. Optimal primal-dual solution can be approximated by simple switching subgradient schemes.

2. Approximations of dual multipliers have natural interpretation

Conclusion

1. Optimal primal-dual solution can be approximated by simple switching subgradient schemes.

2. Approximations of dual multipliers have natural interpretation: relative importance of corresponding constraints during the adjustments process.

Conclusion

1. Optimal primal-dual solution can be approximated by simple switching subgradient schemes.

2. Approximations of dual multipliers have natural interpretation: relative importance of corresponding constraints during the adjustments process.

3. However, it has optimal worst-case efficiency estimate

Conclusion

1. Optimal primal-dual solution can be approximated by simple switching subgradient schemes.

2. Approximations of dual multipliers have natural interpretation: relative importance of corresponding constraints during the adjustments process.

3. However, it has optimal worst-case efficiency estimate even if the dual optimal solution does not exist.

Conclusion

1. Optimal primal-dual solution can be approximated by simple switching subgradient schemes.

2. Approximations of dual multipliers have natural interpretation: relative importance of corresponding constraints during the adjustments process.

3. However, it has optimal worst-case efficiency estimate even if the dual optimal solution does not exist.

4. Many interesting questions (influence of smoothness, strong convexity, etc.)

Conclusion

1. Optimal primal-dual solution can be approximated by simple switching subgradient schemes.

2. Approximations of dual multipliers have natural interpretation: relative importance of corresponding constraints during the adjustments process.

3. However, it has optimal worst-case efficiency estimate even if the dual optimal solution does not exist.

4. Many interesting questions (influence of smoothness, strong convexity, etc.)

THANK YOU FOR YOUR ATTENTION!