Dual methods for functions with bounded variation

Yurii Nesterov, CORE/INMA (UCL)

November 7, 2013, Premolab (Moscow)

Joint work with A.Gasnikov (MIPT, Moscow)

Outline

1 Problem formulation

2 Bounds on the dual solution

3 Problems with bounded variation

4 Modified Gradient Methods

5 Fast Gradient Method

6 Examples

Problem formulation

Problem:

$$(\mathcal{P}) : \qquad f^* \stackrel{\text{def}}{=} \min_{x \in Q} \{ f(x) : \; Ax = b \},$$

Problem formulation

Problem:

$$(\mathcal{P}) : \qquad f^* \stackrel{\mathrm{def}}{=} \min_{x \in Q} \{ f(x) : \; Ax = b \},$$

- $Q \subset X$ is a simple closed convex set,

Problem formulation

Problem:

$$(\mathcal{P}): \qquad f^* \stackrel{\mathrm{def}}{=} \min_{x \in Q} \{ f(x) : \; Ax = b \},$$

- $Q \subset X$ is a simple closed convex set,
- function $f(\cdot)$ is strongly convex on Q:

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

Problem formulation

Problem:

$$(\mathcal{P}): \qquad f^* \stackrel{\mathrm{def}}{=} \min_{x \in Q} \{ f(x) : \; Ax = b \},$$

- $Q \subset X$ is a simple closed convex set,
- function $f(\cdot)$ is strongly convex on Q:

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

- $A : X \to Y^*$ is a linear operator, and $b \in Y^*$.

Problem formulation

Problem:

$$(\mathcal{P}): \qquad f^* \overset{\text{def}}{=} \min_{x \in Q} \{ f(x) : \; Ax = b \},$$

- $Q \subset X$ is a simple closed convex set,
- function $f(\cdot)$ is strongly convex on Q:

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

- $A : X \to Y^*$ is a linear operator, and $b \in Y^*$.

Approximate solution: find $\bar{x} \in Q$ such that

$$f(\bar{x}) - f^* \leq \epsilon_f, \quad \|A\bar{x} - b\| \; \leq \; \epsilon_g.$$

Complexity of problem ($\mathcal{P}$): Black Box model

Complexity of problem ($\mathcal{P}$): Black Box model

Assume for a moment that all norms are Euclidean.

Complexity of problem $(\mathcal{P})$: Black Box model

Assume for a moment that all norms are Euclidean.

- If $A \equiv \emptyset$, $\|\nabla f(x)\|_* \leq L$, and $x \in Q$, then we need
$$O\left(\frac{L^2}{\epsilon_f \; \sigma(f)}\right) \text{ iterations.}$$

Complexity of problem ($\mathcal{P}$): Black Box model

Assume for a moment that all norms are Euclidean.

- If $A \equiv \emptyset$, $\|\nabla f(x)\|_* \leq L$, and $x \in Q$, then we need
$$O\left(\frac{L^2}{\epsilon_f \; \sigma(f)}\right) \text{ iterations.}$$
- If $f \equiv \text{const}$, then we need $O\left(\frac{1}{\epsilon_g} \; \|A\| \operatorname{diam} Q\right)$ iterations.

Complexity of problem $(\mathcal{P})$: Black Box model

Assume for a moment that all norms are Euclidean.

- If $A \equiv \emptyset$, $\|\nabla f(x)\|_* \leq L$, and $x \in Q$, then we need
  $$O\left(\frac{L^2}{\epsilon_f \; \sigma(f)}\right) \text{ iterations.}$$
- If $f \equiv \text{const}$, then we need $O\left(\frac{1}{\epsilon_g} \; \|A\| \; \mathrm{diam}\, Q\right)$ iterations.
  (Minimize the squared residual by FGM).

Complexity of problem $(\mathcal{P})$: Black Box model

Assume for a moment that all norms are Euclidean.

- If $A \equiv \emptyset$, $\|\nabla f(x)\|_* \leq L$, and $x \in Q$, then we need
  $$O\left(\frac{L^2}{\epsilon_f \; \sigma(f)}\right) \text{ iterations.}$$
- If $f \equiv \text{const}$, then we need $O\left(\frac{1}{\epsilon_g} \; \|A\| \operatorname{diam} Q\right)$ iterations.
  (Minimize the squared residual by FGM).
- Define nonsmooth functional constraint
  $$g(x) = \|Ax - b\| \leq 0.$$

Complexity of problem $(\mathcal{P})$: Black Box model

Assume for a moment that all norms are Euclidean.

- If $A \equiv \emptyset$, $\|\nabla f(x)\|_* \leq L$, and $x \in Q$, then we need
$$O\left(\frac{L^2}{\epsilon_f \ \sigma(f)}\right) \text{ iterations.}$$
- If $f \equiv \text{const}$, then we need $O\left(\frac{1}{\epsilon_g} \ \|A\| \operatorname{diam} Q\right)$ iterations.
(Minimize the squared residual by FGM).
- Define nonsmooth functional constraint
$$g(x) = \|Ax - b\| \leq 0.$$
Standard subgradient method ensures
$$O\left(\left[\frac{L}{\epsilon_f} + \frac{\|A\|}{\epsilon_g}\right]^2 \operatorname{diam}^2 Q\right) \text{ iterations.}$$

Dual approach

Dual approach

Define the Lagrangian $\mathcal{L}(x, y) = f(x) + \langle b - Ax, y \rangle$.

Dual approach

Define the Lagrangian $\mathcal{L}(x,y) = f(x) + \langle b - Ax, y\rangle$. The dual problem is as follows:

$$\phi(y) \stackrel{\mathrm{def}}{=} \min_{x\in Q} \mathcal{L}(x,y), \quad \phi^* \stackrel{\mathrm{def}}{=} \sup_{y\in Y} \phi(y).$$

Dual approach

Define the Lagrangian $\mathcal{L}(x, y) = f(x) + \langle b - Ax, y\rangle$. The dual problem is as follows:

$$\phi(y) \stackrel{\mathrm{def}}{=} \min_{x \in Q} \mathcal{L}(x, y), \quad \phi^* \stackrel{\mathrm{def}}{=} \sup_{y \in Y} \phi(y).$$

NB: This is useful only if $\phi(y)$ can be easily computed.

Dual approach

Define the Lagrangian $\mathcal{L}(x,y) = f(x) + \langle b - Ax, y\rangle$. The dual problem is as follows:

$$\phi(y) \stackrel{\text{def}}{=} \min_{x\in Q} \mathcal{L}(x,y), \quad \phi^* \stackrel{\text{def}}{=} \sup_{y\in Y} \phi(y).$$

NB: This is useful only if $\phi(y)$ can be easily computed.

Since f is strongly convex, $\phi(y)$ is well defined for any $y \in Y$:

Dual approach

Define the Lagrangian $\mathcal{L}(x,y) = f(x) + \langle b - Ax, y\rangle$. The dual problem is as follows:

$$\phi(y) \stackrel{\text{def}}{=} \min_{x\in Q} \mathcal{L}(x,y), \quad \phi^* \stackrel{\text{def}}{=} \sup_{y\in Y} \phi(y).$$

NB: This is useful only if $\phi(y)$ can be easily computed.

Since f is strongly convex, $\phi(y)$ is well defined for any $y \in Y$:

$$\phi(y) = \mathcal{L}(x(y), y), \quad \nabla\phi(y) = b - Ax(y), \quad y \in Y,$$

where the point $x(y) \in Q$ is uniquely defined.

Dual approach

Define the Lagrangian $\mathcal{L}(x,y) = f(x) + \langle b - Ax, y \rangle$. The dual problem is as follows:

$$\phi(y) \stackrel{\text{def}}{=} \min_{x \in Q} \mathcal{L}(x,y), \quad \phi^* \stackrel{\text{def}}{=} \sup_{y \in Y} \phi(y).$$

NB: This is useful only if $\phi(y)$ can be easily computed.

Since f is strongly convex, $\phi(y)$ is well defined for any $y \in Y$:

$$\phi(y) = \mathcal{L}(x(y), y), \quad \nabla \phi(y) = b - Ax(y), \quad y \in Y,$$

where the point $x(y) \in Q$ is uniquely defined. Its gradient is Lipschitz continuous with $L(\phi) = \frac{1}{\sigma(f)} \|A\|^2$.

Dual approach

Define the Lagrangian $\mathcal{L}(x,y) = f(x) + \langle b - Ax, y\rangle$. The dual problem is as follows:

$$\phi(y) \stackrel{\text{def}}{=} \min_{x\in Q} \mathcal{L}(x,y), \quad \phi^* \stackrel{\text{def}}{=} \sup_{y\in Y} \phi(y).$$

NB: This is useful only if $\phi(y)$ can be easily computed.

Since f is strongly convex, $\phi(y)$ is well defined for any $y \in Y$:

$$\phi(y) = \mathcal{L}(x(y),y), \quad \nabla\phi(y) = b - Ax(y), \quad y \in Y,$$

where the point $x(y) \in Q$ is uniquely defined. Its gradient is Lipschitz continuous with $L(\phi) = \frac{1}{\sigma(f)}\|A\|^2$.

But:

Dual approach

Define the Lagrangian $\mathcal{L}(x,y) = f(x) + \langle b - Ax, y\rangle$. The dual problem is as follows:

$$\phi(y) \stackrel{\text{def}}{=} \min_{x\in Q} \mathcal{L}(x,y), \quad \phi^* \stackrel{\text{def}}{=} \sup_{y\in Y} \phi(y).$$

NB: This is useful only if $\phi(y)$ can be easily computed.

Since f is strongly convex, $\phi(y)$ is well defined for any $y \in Y$:

$$\phi(y) = \mathcal{L}(x(y),y), \quad \nabla\phi(y) = b - Ax(y), \quad y \in Y,$$

where the point $x(y) \in Q$ is uniquely defined. Its gradient is Lipschitz continuous with $L(\phi) = \frac{1}{\sigma(f)}\|A\|^2$.

But: the standard complexity bounds of GMs depend on y^*:

Dual approach

Define the Lagrangian $\mathcal{L}(x,y) = f(x) + \langle b - Ax, y\rangle$. The dual problem is as follows:

$$\phi(y) \stackrel{\mathrm{def}}{=} \min_{x\in Q} \mathcal{L}(x,y), \quad \phi^* \stackrel{\mathrm{def}}{=} \sup_{y\in Y} \phi(y).$$

NB: This is useful only if $\phi(y)$ can be easily computed.

Since f is strongly convex, $\phi(y)$ is well defined for any $y \in Y$:

$$\phi(y) = \mathcal{L}(x(y), y), \quad \nabla\phi(y) = b - Ax(y), \quad y \in Y,$$

where the point $x(y) \in Q$ is uniquely defined. Its gradient is Lipschitz continuous with $L(\phi) = \frac{1}{\sigma(f)}\|A\|^2$.

But: the standard complexity bounds of GMs depend on y^*: Choosing $y_0 = 0$,

Dual approach

Define the Lagrangian $\mathcal{L}(x,y) = f(x) + \langle b - Ax, y\rangle$. The dual problem is as follows:

$$\phi(y) \stackrel{\mathrm{def}}{=} \min_{x\in Q} \mathcal{L}(x,y), \quad \phi^* \stackrel{\mathrm{def}}{=} \sup_{y\in Y} \phi(y).$$

NB: This is useful only if $\phi(y)$ can be easily computed.

Since f is strongly convex, $\phi(y)$ is well defined for any $y \in Y$:

$$\phi(y) = \mathcal{L}(x(y),y), \quad \nabla\phi(y) = b - Ax(y), \quad y \in Y,$$

where the point $x(y) \in Q$ is uniquely defined. Its gradient is Lipschitz continuous with $L(\phi) = \frac{1}{\sigma(f)}\|A\|^2$.

But: the standard complexity bounds of GMs depend on y^*: Choosing $y_0 = 0$, we have $\phi(y^*) - \phi(y_k) \le \frac{4L(\phi)\|y^*\|^2}{(k+1)^2}$.

Dual approach

Define the Lagrangian $\mathcal{L}(x,y) = f(x) + \langle b - Ax, y\rangle$. The dual problem is as follows:

$$\phi(y) \overset{\text{def}}{=} \min_{x \in Q} \mathcal{L}(x,y), \quad \phi^* \overset{\text{def}}{=} \sup_{y \in Y} \phi(y).$$

NB: This is useful only if $\phi(y)$ can be easily computed.

Since f is strongly convex, $\phi(y)$ is well defined for any $y \in Y$:

$$\phi(y) = \mathcal{L}(x(y), y), \quad \nabla\phi(y) = b - Ax(y), \quad y \in Y,$$

where the point $x(y) \in Q$ is uniquely defined. Its gradient is Lipschitz continuous with $L(\phi) = \frac{1}{\sigma(f)}\|A\|^2$.

But: the standard complexity bounds of GMs depend on y^*: Choosing $y_0 = 0$, we have $\phi(y^*) - \phi(y_k) \le \frac{4L(\phi)\|y^*\|^2}{(k+1)^2}$.

NB: $\|y^*\|$ can be big!

Example

Example

Let $E = R^2$, and $\|x\| \equiv \|x\|_2$.

Example

Let $E = R^2$, and $\|x\| \equiv \|x\|_2$. Consider the problem

$$\min_x \left\{ f(x) = \tfrac{1}{2}\|x - e_2\|^2 : \; x^{(1)} = 1, \; \|x\| \leq 1 \right\}.$$

Example

Let $E = R^2$, and $\|x\| \equiv \|x\|_2$. Consider the problem

$$\min_x \left\{ f(x) = \tfrac{1}{2}\|x - e_2\|^2 : \; x^{(1)} = 1, \; \|x\| \le 1 \right\}.$$

Defining the Lagrangian

$$\mathcal{L}(x,y) = \tfrac{1}{2}\|x - e_2\|^2 + y(1 - x^{(1)}) \;=\; \tfrac{1}{2}\|x - e_2 - ye_1\|^2 + y - \tfrac{1}{2}y^2,$$

we get $x(y) = \hat{x}(y)/\|\hat{x}(y)\|$, where $\hat{x}(y) = e_2 + ye_1$.

Example

Let $E = R^2$, and $\|x\| \equiv \|x\|_2$. Consider the problem

$$\min_x \left\{ f(x) = \tfrac{1}{2}\|x - e_2\|^2 : \; x^{(1)} = 1, \; \|x\| \leq 1 \right\}.$$

Defining the Lagrangian

$$\mathcal{L}(x,y) = \tfrac{1}{2}\|x - e_2\|^2 + y(1 - x^{(1)}) \;=\; \tfrac{1}{2}\|x - e_2 - ye_1\|^2 + y - \tfrac{1}{2}y^2,$$

we get $x(y) = \hat{x}(y)/\|\hat{x}(y)\|$, where $\hat{x}(y) = e_2 + ye_1$.

Thus, $\phi(y) = \frac{1}{2}(\|\hat{x}(y)\| - 1)^2 + y - \frac{1}{2}y^2 \;=\; 1 - \frac{1}{y + \sqrt{1+y^2}}$.

Example

Let $E = R^2$, and $\|x\| \equiv \|x\|_2$. Consider the problem

$$\min_x \left\{ f(x) = \tfrac{1}{2}\|x - e_2\|^2 : \; x^{(1)} = 1, \; \|x\| \leq 1 \right\}.$$

Defining the Lagrangian

$$\mathcal{L}(x,y) = \tfrac{1}{2}\|x - e_2\|^2 + y(1 - x^{(1)}) \; = \; \tfrac{1}{2}\|x - e_2 - ye_1\|^2 + y - \tfrac{1}{2}y^2,$$

we get $x(y) = \hat{x}(y)/\|\hat{x}(y)\|$, where $\hat{x}(y) = e_2 + ye_1$.

Thus, $\phi(y) = \frac{1}{2}(\|\hat{x}(y)\| - 1)^2 + y - \frac{1}{2}y^2 \; = \; 1 - \frac{1}{y + \sqrt{1+y^2}}$.

- No duality gap and $\phi(y) \to f^* = 1$ as $y \to \infty$.

Example

Let $E = R^2$, and $\|x\| \equiv \|x\|_2$. Consider the problem

$$\min_x \left\{ f(x) = \tfrac{1}{2}\|x - e_2\|^2 : \; x^{(1)} = 1, \; \|x\| \leq 1 \right\}.$$

Defining the Lagrangian

$$\mathcal{L}(x, y) = \tfrac{1}{2}\|x - e_2\|^2 + y(1 - x^{(1)}) \; = \; \tfrac{1}{2}\|x - e_2 - ye_1\|^2 + y - \tfrac{1}{2}y^2,$$

we get $x(y) = \hat{x}(y)/\|\hat{x}(y)\|$, where $\hat{x}(y) = e_2 + ye_1$.

Thus, $\phi(y) = \frac{1}{2}(\|\hat{x}(y)\| - 1)^2 + y - \frac{1}{2}y^2 \; = \; 1 - \frac{1}{y + \sqrt{1+y^2}}$.

- No duality gap and $\phi(y) \to f^* = 1$ as $y \to \infty$.
- $x(y) \to x^* = e_1$ as $y \to \infty$.

Example

Let $E = R^2$, and $\|x\| \equiv \|x\|_2$. Consider the problem

$$\min_x \left\{ f(x) = \tfrac{1}{2}\|x - e_2\|^2 : \; x^{(1)} = 1, \; \|x\| \le 1 \right\}.$$

Defining the Lagrangian

$$\mathcal{L}(x,y) = \tfrac{1}{2}\|x - e_2\|^2 + y(1 - x^{(1)}) \; = \; \tfrac{1}{2}\|x - e_2 - ye_1\|^2 + y - \tfrac{1}{2}y^2,$$

we get $x(y) = \hat{x}(y)/\|\hat{x}(y)\|$, where $\hat{x}(y) = e_2 + ye_1$.

Thus, $\phi(y) = \frac{1}{2}(\|\hat{x}(y)\| - 1)^2 + y - \frac{1}{2}y^2 \; = \; 1 - \frac{1}{y+\sqrt{1+y^2}}$.

- No duality gap and $\phi(y) \to f^* = 1$ as $y \to \infty$.
- $x(y) \to x^* = e_1$ as $y \to \infty$.
- Optimal solution does not exist.

Example

Let $E = R^2$, and $\|x\| \equiv \|x\|_2$. Consider the problem

$$\min_x \left\{ f(x) = \tfrac{1}{2}\|x - e_2\|^2 : \; x^{(1)} = 1, \; \|x\| \leq 1 \right\}.$$

Defining the Lagrangian

$\mathcal{L}(x,y) = \frac{1}{2}\|x - e_2\|^2 + y(1 - x^{(1)}) \; = \; \frac{1}{2}\|x - e_2 - ye_1\|^2 + y - \frac{1}{2}y^2,$

we get $x(y) = \hat{x}(y)/\|\hat{x}(y)\|$, where $\hat{x}(y) = e_2 + ye_1$.

Thus, $\phi(y) = \frac{1}{2}(\|\hat{x}(y)\| - 1)^2 + y - \frac{1}{2}y^2 \; = \; 1 - \frac{1}{y+\sqrt{1+y^2}}$.

- No duality gap and $\phi(y) \to f^* = 1$ as $y \to \infty$.
- $x(y) \to x^* = e_1$ as $y \to \infty$.
- Optimal solution does not exist. Rate of convergence of the standard dual GMs = ?

Bounding the dual solution

Bounding the dual solution

Assumption 1: $\exists \bar{x}$ and a radius $\rho > 0$ such that

Bounding the dual solution

Assumption 1: $\exists \bar{x}$ and a radius $\rho > 0$ such that

$$A\bar{x} = b \text{ and } B(\bar{x}, \rho) \subseteq Q.$$

Bounding the dual solution

Assumption 1: $\exists \bar{x}$ and a radius $\rho > 0$ such that

$$A\bar{x} = b \text{ and } B(\bar{x}, \rho) \subseteq Q.$$

Lemma. $\|A^T y^*\| \leq (1 + \frac{1}{\rho} D)\|\nabla f(x^*)\|_*$, where $D = \operatorname{diam} Q$.

Bounding the dual solution

Assumption 1: $\exists \bar{x}$ and a radius $\rho > 0$ such that

$$A\bar{x} = b \text{ and } B(\bar{x}, \rho) \subseteq Q.$$

Lemma. $\|A^T y^*\| \leq (1 + \frac{1}{\rho} D)\|\nabla f(x^*)\|_*$, where $D = \operatorname{diam} Q$.

Proof: since $\langle \nabla f(x^*) - A^T y^*, x - x^* \rangle \geq 0$,

Bounding the dual solution

Assumption 1: $\exists \bar{x}$ and a radius $\rho > 0$ such that

$$A\bar{x} = b \text{ and } B(\bar{x}, \rho) \subseteq Q.$$

Lemma. $\|A^T y^*\| \leq (1 + \frac{1}{\rho} D)\|\nabla f(x^*)\|_*$, where $D = \operatorname{diam} Q$.

Proof: since $\langle \nabla f(x^*) - A^T y^*, x - x^* \rangle \geq 0$, for all $x \in Q$ we have

$$0 \leq \langle \nabla f(x^*) - A^T y^*, \bar{x} - x^* \rangle + \langle \nabla f(x^*) - A^T y^*, x - \bar{x} \rangle$$

Bounding the dual solution

Assumption 1: $\exists \bar{x}$ and a radius $\rho > 0$ such that

$$A\bar{x} = b \text{ and } B(\bar{x}, \rho) \subseteq Q.$$

Lemma. $\|A^T y^*\| \leq (1 + \frac{1}{\rho} D) \|\nabla f(x^*)\|_*$, where $D = \operatorname{diam} Q$.

Proof: since $\langle \nabla f(x^*) - A^T y^*, x - x^* \rangle \geq 0$, for all $x \in Q$ we have

$$\begin{aligned} 0 &\leq \langle \nabla f(x^*) - A^T y^*, \bar{x} - x^* \rangle + \langle \nabla f(x^*) - A^T y^*, x - \bar{x} \rangle \\ &= \langle \nabla f(x^*), \bar{x} - x^* \rangle + \langle \nabla f(x^*) - A^T y^*, x - \bar{x} \rangle. \end{aligned}$$

Bounding the dual solution

Assumption 1: $\exists \bar{x}$ and a radius $\rho > 0$ such that

$$A\bar{x} = b \text{ and } B(\bar{x}, \rho) \subseteq Q.$$

Lemma. $\|A^T y^*\| \leq (1 + \frac{1}{\rho} D)\|\nabla f(x^*)\|_*$, where $D = \operatorname{diam} Q$.

Proof: since $\langle \nabla f(x^*) - A^T y^*, x - x^* \rangle \geq 0$, for all $x \in Q$ we have

$$\begin{array}{rcl} 0 & \leq & \langle \nabla f(x^*) - A^T y^*, \bar{x} - x^* \rangle + \langle \nabla f(x^*) - A^T y^*, x - \bar{x} \rangle \\ & = & \langle \nabla f(x^*), \bar{x} - x^* \rangle + \langle \nabla f(x^*) - A^T y^*, x - \bar{x} \rangle. \end{array}$$

Restricting ourselves to $x \in B(\bar{x}, \rho)$, we obtain

$$\rho \|\nabla f(x^*) - A^T y^*\|_* \;\; \leq \;\; \langle \nabla f(x^*), \bar{x} - x^* \rangle \;\; \leq \;\; \|\nabla f(x^*)\|_* \cdot D.$$

Bounding the dual solution

Assumption 1: $\exists \bar{x}$ and a radius $\rho > 0$ such that

$$A\bar{x} = b \text{ and } B(\bar{x}, \rho) \subseteq Q.$$

Lemma. $\|A^T y^*\| \leq (1 + \frac{1}{\rho} D)\|\nabla f(x^*)\|_*$, where $D = \operatorname{diam} Q$.

Proof: since $\langle \nabla f(x^*) - A^T y^*, x - x^* \rangle \geq 0$, for all $x \in Q$ we have

$$\begin{aligned} 0 &\leq \langle \nabla f(x^*) - A^T y^*, \bar{x} - x^* \rangle + \langle \nabla f(x^*) - A^T y^*, x - \bar{x} \rangle \\ &= \langle \nabla f(x^*), \bar{x} - x^* \rangle + \langle \nabla f(x^*) - A^T y^*, x - \bar{x} \rangle. \end{aligned}$$

Restricting ourselves to $x \in B(\bar{x}, \rho)$, we obtain

$$\rho \|\nabla f(x^*) - A^T y^*\|_* \leq \langle \nabla f(x^*), \bar{x} - x^* \rangle \leq \|\nabla f(x^*)\|_* \cdot D.$$

Hence, $\|\nabla f(x^*)\|_* \cdot D \geq \rho \|A^T y^*\|_* - \rho \|\nabla f(x^*)\|_*$. □

Bounding the dual solution

Assumption 1: $\exists \bar{x}$ and a radius $\rho > 0$ such that

$$A\bar{x} = b \text{ and } B(\bar{x}, \rho) \subseteq Q.$$

Lemma. $\|A^T y^*\| \leq (1 + \frac{1}{\rho} D)\|\nabla f(x^*)\|_*$, where $D = \operatorname{diam} Q$.

Proof: since $\langle \nabla f(x^*) - A^T y^*, x - x^* \rangle \geq 0$, for all $x \in Q$ we have

$$\begin{array}{rcl} 0 & \leq & \langle \nabla f(x^*) - A^T y^*, \bar{x} - x^* \rangle + \langle \nabla f(x^*) - A^T y^*, x - \bar{x} \rangle \\ & = & \langle \nabla f(x^*), \bar{x} - x^* \rangle + \langle \nabla f(x^*) - A^T y^*, x - \bar{x} \rangle. \end{array}$$

Restricting ourselves to $x \in B(\bar{x}, \rho)$, we obtain

$$\rho \|\nabla f(x^*) - A^T y^*\|_* \;\leq\; \langle \nabla f(x^*), \bar{x} - x^* \rangle \;\leq\; \|\nabla f(x^*)\|_* \cdot D.$$

Hence, $\|\nabla f(x^*)\|_* \cdot D \geq \rho \|A^T y^*\|_* - \rho \|\nabla f(x^*)\|_*$. □

NB: We need $\|\nabla f(x)\|_* \leq \text{const}$.

Bounding the dual solution

Assumption 1: $\exists \bar{x}$ and a radius $\rho > 0$ such that

$$A\bar{x} = b \text{ and } B(\bar{x}, \rho) \subseteq Q.$$

Lemma. $\|A^T y^*\| \leq (1 + \frac{1}{\rho} D)\|\nabla f(x^*)\|_*$, where $D = \mathrm{diam}\, Q$.

Proof: since $\langle \nabla f(x^*) - A^T y^*, x - x^* \rangle \geq 0$, for all $x \in Q$ we have

$$\begin{array}{rcl} 0 & \leq & \langle \nabla f(x^*) - A^T y^*, \bar{x} - x^* \rangle + \langle \nabla f(x^*) - A^T y^*, x - \bar{x} \rangle \\ & = & \langle \nabla f(x^*), \bar{x} - x^* \rangle + \langle \nabla f(x^*) - A^T y^*, x - \bar{x} \rangle. \end{array}$$

Restricting ourselves to $x \in B(\bar{x}, \rho)$, we obtain

$$\rho \|\nabla f(x^*) - A^T y^*\|_* \;\leq\; \langle \nabla f(x^*), \bar{x} - x^* \rangle \;\leq\; \|\nabla f(x^*)\|_* \cdot D.$$

Hence, $\|\nabla f(x^*)\|_* \cdot D \geq \rho \|A^T y^*\|_* - \rho \|\nabla f(x^*)\|_*$. □

NB: We need $\|\nabla f(x)\|_* \leq$ const. (This may not happen.)

Problems with bounded variation

Problems with bounded variation

Assume that problem $(\mathcal{P})$ is solvable.

Problems with bounded variation

Assume that problem $(\mathcal{P})$ is solvable. Define its *outer variation*

$$\mathrm{Out}(\mathcal{P}) \;=\; \min_{x \in Q}\{f(x) : \; Ax = b\} - \min_{x \in Q} f(x).$$

Problems with bounded variation

Assume that problem $(\mathcal{P})$ is solvable. Define its *outer variation*

$$\mathrm{Out}(\mathcal{P}) \;=\; \min_{x \in Q}\{ f(x) : \; Ax = b \} \;-\; \min_{x \in Q} f(x).$$

Since f is strongly convex, this value is finite.

Problems with bounded variation

Assume that problem $(\mathcal{P})$ is solvable. Define its *outer variation*

$$\text{Out}(\mathcal{P}) \;\; = \;\; \min_{x \in Q} \{ f(x) : \; Ax = b \} \; - \; \min_{x \in Q} f(x).$$

Since f is strongly convex, this value is finite.
On the other hand, $\phi(0) = \min\limits_{x \in Q} f(x) \stackrel{\text{def}}{=} f(x_*)$, and

$$\phi(y) \;\; \leq \;\; \mathcal{L}(x^*, y) \; = \; f^*, \quad y \in Y, \tag{1}$$

Problems with bounded variation

Assume that problem $(\mathcal{P})$ is solvable. Define its *outer variation*

$$\mathrm{Out}(\mathcal{P}) \;=\; \min_{x \in Q}\{f(x) : \; Ax = b\} \;-\; \min_{x \in Q} f(x).$$

Since f is strongly convex, this value is finite.
On the other hand, $\phi(0) = \min\limits_{x \in Q} f(x) \overset{\mathrm{def}}{=} f(x_*)$, and

$$\phi(y) \;\leq\; \mathcal{L}(x^*, y) \;=\; f^*, \quad y \in Y, \tag{1}$$

Thus, function ϕ has a bounded *central* variation on Y:

$$\phi^* - \phi(0) \;=\; \mathrm{Out}(\mathcal{P}) \;<\; +\infty. \tag{2}$$

Problems with bounded variation

Assume that problem $(\mathcal{P})$ is solvable. Define its *outer variation*

$$\mathrm{Out}(\mathcal{P}) \;=\; \min_{x \in Q} \{ f(x) : \; Ax = b \} \;-\; \min_{x \in Q} f(x).$$

Since f is strongly convex, this value is finite.
On the other hand, $\phi(0) = \min\limits_{x \in Q} f(x) \stackrel{\mathrm{def}}{=} f(x_*)$, and

$$\phi(y) \;\leq\; \mathcal{L}(x^*, y) \;=\; f^*, \quad y \in Y, \tag{1}$$

Thus, function ϕ has a bounded *central* variation on Y:

$$\phi^* - \phi(0) \;=\; \mathrm{Out}(\mathcal{P}) \;<\; +\infty. \tag{2}$$

We study numerical schemes for maximizing dual functions satisfying assumption (2).

Termination criterion

Termination criterion

For given $\epsilon_f, \epsilon_g > 0$, we are going to generate $\bar{y} \in Y$:

Termination criterion

For given $\epsilon_f, \epsilon_g > 0$, we are going to generate $\bar{y} \in Y$:

$$\text{a)} \quad \|\nabla \phi(\bar{y})\|_* \;\leq\; \epsilon_g,$$

Termination criterion

For given $\epsilon_f, \epsilon_g > 0$, we are going to generate $\bar{y} \in Y$:

$$\begin{array}{lrcl} \text{a)} & \|\nabla \phi(\bar{y})\|_* & \leq & \epsilon_g, \\ \text{b)} & \langle \nabla \phi(\bar{y}), \bar{y} \rangle & \geq & -\epsilon_f. \end{array} \qquad (3)$$

Termination criterion

For given $\epsilon_f, \epsilon_g > 0$, we are going to generate $\bar{y} \in Y$:

$$
\begin{aligned}
\text{a)} \quad \|\nabla \phi(\bar{y})\|_* \;&\leq\; \epsilon_g, \\
\text{b)} \quad \langle \nabla \phi(\bar{y}), \bar{y} \rangle \;&\geq\; -\epsilon_f.
\end{aligned} \tag{3}
$$

Lemma. Let point $\bar{y} \in Y$ satisfy conditions (3).

Termination criterion

For given $\epsilon_f, \epsilon_g > 0$, we are going to generate $\bar{y} \in Y$:

$$\begin{array}{rrcl} \text{a)} & \|\nabla \phi(\bar{y})\|_* & \leq & \epsilon_g, \\ \text{b)} & \langle \nabla \phi(\bar{y}), \bar{y} \rangle & \geq & -\epsilon_f. \end{array} \qquad (3)$$

Lemma. Let point $\bar{y} \in Y$ satisfy conditions (3). Then

$$x(\bar{y}) \in Q, \quad \|Ax(\bar{y}) - b\| \leq \epsilon_g, \quad f(x(\bar{y})) \leq f^* + \epsilon_f.$$

Termination criterion

For given $\epsilon_f, \epsilon_g > 0$, we are going to generate $\bar{y} \in Y$:

$$\begin{array}{rrcl} \text{a)} & \|\nabla \phi(\bar{y})\|_* & \leq & \epsilon_g, \\ \text{b)} & \langle \nabla \phi(\bar{y}), \bar{y} \rangle & \geq & -\epsilon_f. \end{array} \tag{3}$$

Lemma. Let point $\bar{y} \in Y$ satisfy conditions (3). Then

$$x(\bar{y}) \in Q, \quad \|Ax(\bar{y}) - b\| \leq \epsilon_g, \quad f(x(\bar{y})) \leq f^* + \epsilon_f.$$

Proof. Indeed, the point $x(\bar{y})$ belongs to Q by definition,

Termination criterion

For given $\epsilon_f, \epsilon_g > 0$, we are going to generate $\bar{y} \in Y$:

$$\begin{array}{lrcl} \text{a)} & \|\nabla \phi(\bar{y})\|_* & \leq & \epsilon_g, \\ \text{b)} & \langle \nabla \phi(\bar{y}), \bar{y} \rangle & \geq & -\epsilon_f. \end{array} \quad (3)$$

Lemma. Let point $\bar{y} \in Y$ satisfy conditions (3). Then

$$x(\bar{y}) \in Q, \quad \|Ax(\bar{y}) - b\| \leq \epsilon_g, \quad f(x(\bar{y})) \leq f^* + \epsilon_f.$$

Proof. Indeed, the point $x(\bar{y})$ belongs to Q by definition, and $\nabla \phi(\bar{y}) = b - Ax(\bar{y})$.

Termination criterion

For given $\epsilon_f, \epsilon_g > 0$, we are going to generate $\bar{y} \in Y$:

$$\begin{array}{ll} \text{a)} & \|\nabla \phi(\bar{y})\|_* \;\leq\; \epsilon_g, \\ & \\ \text{b)} & \langle \nabla \phi(\bar{y}), \bar{y} \rangle \;\geq\; -\epsilon_f. \end{array} \qquad (3)$$

Lemma. Let point $\bar{y} \in Y$ satisfy conditions (3). Then

$$x(\bar{y}) \in Q, \quad \|Ax(\bar{y}) - b\| \leq \epsilon_g, \quad f(x(\bar{y})) \leq f^* + \epsilon_f.$$

Proof. Indeed, the point $x(\bar{y})$ belongs to Q by definition, and $\nabla \phi(\bar{y}) = b - Ax(\bar{y})$. Moreover,

$$f(x(\bar{y})) + \langle b - Ax(\bar{y}), \bar{y} \rangle$$

Termination criterion

For given $\epsilon_f, \epsilon_g > 0$, we are going to generate $\bar{y} \in Y$:

$$\begin{aligned} \text{a)} \quad \|\nabla \phi(\bar{y})\|_* \;&\leq\; \epsilon_g, \\ \text{b)} \quad \langle \nabla \phi(\bar{y}), \bar{y} \rangle \;&\geq\; -\epsilon_f. \end{aligned} \tag{3}$$

Lemma. Let point $\bar{y} \in Y$ satisfy conditions (3). Then

$$x(\bar{y}) \in Q, \quad \|Ax(\bar{y}) - b\| \leq \epsilon_g, \quad f(x(\bar{y})) \leq f^* + \epsilon_f.$$

Proof. Indeed, the point $x(\bar{y})$ belongs to Q by definition, and $\nabla \phi(\bar{y}) = b - Ax(\bar{y})$. Moreover,

$$f(x(\bar{y})) + \langle b - Ax(\bar{y}), \bar{y} \rangle \leq f(x^*) + \langle b - Ax^*, \bar{y} \rangle$$

Termination criterion

For given $\epsilon_f, \epsilon_g > 0$, we are going to generate $\bar{y} \in Y$:

$$\begin{array}{ll} \text{a)} & \|\nabla \phi(\bar{y})\|_* \;\leq\; \epsilon_g, \\[2ex] \text{b)} & \langle \nabla \phi(\bar{y}), \bar{y} \rangle \;\geq\; -\epsilon_f. \end{array} \qquad (3)$$

Lemma. Let point $\bar{y} \in Y$ satisfy conditions (3). Then

$$x(\bar{y}) \in Q, \quad \|Ax(\bar{y}) - b\| \leq \epsilon_g, \quad f(x(\bar{y})) \leq f^* + \epsilon_f.$$

Proof. Indeed, the point $x(\bar{y})$ belongs to Q by definition, and $\nabla \phi(\bar{y}) = b - Ax(\bar{y})$. Moreover,

$$f(x(\bar{y})) + \langle b - Ax(\bar{y}), \bar{y} \rangle \leq f(x^*) + \langle b - Ax^*, \bar{y} \rangle \;=\; f^*.$$

Termination criterion

For given $\epsilon_f, \epsilon_g > 0$, we are going to generate $\bar{y} \in Y$:

$$\begin{aligned} \text{a)} \quad \|\nabla\phi(\bar{y})\|_* &\leq \epsilon_g, \\ \text{b)} \quad \langle \nabla\phi(\bar{y}), \bar{y}\rangle &\geq -\epsilon_f. \end{aligned} \tag{3}$$

Lemma. Let point $\bar{y} \in Y$ satisfy conditions (3). Then

$$x(\bar{y}) \in Q, \quad \|Ax(\bar{y}) - b\| \leq \epsilon_g, \quad f(x(\bar{y})) \leq f^* + \epsilon_f.$$

Proof. Indeed, the point $x(\bar{y})$ belongs to Q by definition, and $\nabla\phi(\bar{y}) = b - Ax(\bar{y})$. Moreover,

$$f(x(\bar{y})) + \langle b - Ax(\bar{y}), \bar{y}\rangle \leq f(x^*) + \langle b - Ax^*, \bar{y}\rangle \; = \; f^*.$$

Hence, $f(x(\bar{y})) \leq f^* - \langle \nabla\phi(\bar{y}), \bar{y}\rangle$

Termination criterion

For given $\epsilon_f, \epsilon_g > 0$, we are going to generate $\bar{y} \in Y$:

$$\begin{aligned} \text{a)} \quad \|\nabla \phi(\bar{y})\|_* \;&\leq\; \epsilon_g, \\ \text{b)} \quad \langle \nabla \phi(\bar{y}), \bar{y} \rangle \;&\geq\; -\epsilon_f. \end{aligned} \tag{3}$$

Lemma. Let point $\bar{y} \in Y$ satisfy conditions (3). Then

$$x(\bar{y}) \in Q, \quad \|Ax(\bar{y}) - b\| \leq \epsilon_g, \quad f(x(\bar{y})) \leq f^* + \epsilon_f.$$

Proof. Indeed, the point $x(\bar{y})$ belongs to Q by definition, and $\nabla \phi(\bar{y}) = b - Ax(\bar{y})$. Moreover,

$$f(x(\bar{y})) + \langle b - Ax(\bar{y}), \bar{y} \rangle \leq f(x^*) + \langle b - Ax^*, \bar{y} \rangle \;=\; f^*.$$

Hence, $f(x(\bar{y})) \leq f^* - \langle \nabla \phi(\bar{y}), \bar{y} \rangle \overset{(3)_b}{\leq} f^* + \epsilon_f$. □

Range of accuracy for the norm of the gradient

Range of accuracy for the norm of the gradient

Conditions (3) with any $\epsilon_f \geq 0$ and $\epsilon_g \geq \|Ax_* - b\|$ are satisfied by $\bar{y} = 0$.

Range of accuracy for the norm of the gradient

Conditions (3) with any $\epsilon_f \geq 0$ and $\epsilon_g \geq \|Ax_* - b\|$ are satisfied by $\bar{y} = 0$.

Therefore, we always assume that

$$\epsilon_g^2 \quad \leq \quad \|Ax_* - b\|^2$$

Range of accuracy for the norm of the gradient

Conditions (3) with any $\epsilon_f \geq 0$ and $\epsilon_g \geq \|Ax_* - b\|$ are satisfied by $\bar{y} = 0$.

Therefore, we always assume that

$$\epsilon_g^2 \;\leq\; \|Ax_* - b\|^2 \;=\; \|A(x_* - x^*)\|^2$$

Range of accuracy for the norm of the gradient

Conditions (3) with any $\epsilon_f \geq 0$ and $\epsilon_g \geq \|Ax_* - b\|$ are satisfied by $\bar{y} = 0$.

Therefore, we always assume that

$$\epsilon_g^2 \;\leq\; \|Ax_* - b\|^2 \;=\; \|A(x_* - x^*)\|^2 \;\leq\; \|A\|^2 \, \|x_* - x^*\|^2$$

Range of accuracy for the norm of the gradient

Conditions (3) with any $\epsilon_f \geq 0$ and $\epsilon_g \geq \|Ax_* - b\|$ are satisfied by $\bar{y} = 0$.

Therefore, we always assume that

$$\begin{aligned} \epsilon_g^2 &\leq \|Ax_* - b\|^2 = \|A(x_* - x^*)\|^2 \leq \|A\|^2 \, \|x_* - x^*\|^2 \\ &\leq \tfrac{2}{\sigma(f)} \|A\|^2 (f(x^*) - f(x_*)) \end{aligned}$$

Range of accuracy for the norm of the gradient

Conditions (3) with any $\epsilon_f \geq 0$ and $\epsilon_g \geq \|Ax_* - b\|$ are satisfied by $\bar{y} = 0$.

Therefore, we always assume that

$$\begin{aligned} \epsilon_g^2 &\leq \|Ax_* - b\|^2 = \|A(x_* - x^*)\|^2 \leq \|A\|^2 \|x_* - x^*\|^2 \\ &\leq \tfrac{2}{\sigma(f)} \|A\|^2 (f(x^*) - f(x_*)) = 2\, L(\phi)\, \mathrm{Out}(\mathcal{P}). \end{aligned}$$

Modified Gradient Method

Modified Gradient Method

Since ϕ has Lipschitz continuous gradients, we can maximize it by a version of GM.

Modified Gradient Method

Since ϕ has Lipschitz continuous gradients, we can maximize it by a version of GM.

Choose starting point $y_0 = 0$. Note that the norms below are arbitrary.

Modified Gradient Method

Since ϕ has Lipschitz continuous gradients, we can maximize it by a version of GM.

Choose starting point $y_0 = 0$. Note that the norms below are arbitrary.

Modified Gradient Method

Modified Gradient Method

Since ϕ has Lipschitz continuous gradients, we can maximize it by a version of GM.
Choose starting point $y_0 = 0$. Note that the norms below are arbitrary.

Modified Gradient Method

1. Choose $y'_k = \arg\max\limits_{y} \left[\langle \nabla \phi(y_k), y - y_k \rangle - \frac{1}{2} L(\phi) \| y - y_k \|^2 \right]$.

Modified Gradient Method

Since ϕ has Lipschitz continuous gradients, we can maximize it by a version of GM.
Choose starting point $y_0 = 0$. Note that the norms below are arbitrary.

Modified Gradient Method

1. Choose $y'_k = \arg\max_y \left[\langle \nabla \phi(y_k), y - y_k \rangle - \frac{1}{2} L(\phi) \| y - y_k \|^2 \right]$.

2. Define $y_{k+1} = t_k y'_k$, where $t_k \in (0, 1]$ is such that

Modified Gradient Method

Since ϕ has Lipschitz continuous gradients, we can maximize it by a version of GM.
Choose starting point $y_0 = 0$. Note that the norms below are arbitrary.

Modified Gradient Method

1. Choose $y'_k = \arg\max\limits_y \left[\langle \nabla\phi(y_k), y - y_k\rangle - \frac{1}{2}L(\phi)\|y - y_k\|^2\right]$.

2. Define $y_{k+1} = t_k y'_k$, where $t_k \in (0, 1]$ is such that

$$\phi(y_{k+1}) \geq \phi(y'_k), \quad \langle \nabla\phi(y_{k+1}), y_{k+1}\rangle \geq -\epsilon_f.$$

Modified Gradient Method

Since ϕ has Lipschitz continuous gradients, we can maximize it by a version of GM.
Choose starting point $y_0 = 0$. Note that the norms below are arbitrary.

Modified Gradient Method

1. Choose $y_k' = \arg\max\limits_y \left[\langle \nabla\phi(y_k), y - y_k \rangle - \frac{1}{2} L(\phi) \| y - y_k \|^2 \right]$.

2. Define $y_{k+1} = t_k y_k'$, where $t_k \in (0,1]$ is such that

$$\phi(y_{k+1}) \geq \phi(y_k'), \quad \langle \nabla\phi(y_{k+1}), y_{k+1} \rangle \geq -\epsilon_f.$$

Conditions of Item 2 can be satisfied by solving 1D-maximization problem $\max\limits_{t \in [0,1]} \phi(t y_k')$.

Convergence of GM

Convergence of GM

Theorem.

Convergence of GM

Theorem. GM is well defined. For any $k \geq 0$ we have

Convergence of GM

Theorem. GM is well defined. For any $k \geq 0$ we have

$$\phi(y_{k+1}) - \phi(y_k) \quad \geq \quad \frac{1}{2L(\phi)} \|\nabla \phi(y_k)\|_*^2. \qquad (4)$$

Convergence of GM

Theorem. GM is well defined. For any $k \geq 0$ we have

$$\phi(y_{k+1}) - \phi(y_k) \quad \geq \quad \tfrac{1}{2L(\phi)} \| \nabla \phi(y_k) \|_*^2. \tag{4}$$

Consequently, $\phi(y_k) \geq \phi(0)$ for all $k \geq 0$,

Convergence of GM

Theorem. GM is well defined. For any $k \geq 0$ we have

$$\phi(y_{k+1}) - \phi(y_k) \quad \geq \quad \frac{1}{2L(\phi)} \|\nabla \phi(y_k)\|_*^2. \qquad (4)$$

Consequently, $\phi(y_k) \geq \phi(0)$ for all $k \geq 0$, and

$$\sum_{k=0}^{N} \|\nabla \phi(y_k)\|_*^2$$

Convergence of GM

Theorem. GM is well defined. For any $k \geq 0$ we have

$$\phi(y_{k+1}) - \phi(y_k) \quad \geq \quad \frac{1}{2L(\phi)} \|\nabla \phi(y_k)\|_*^2. \qquad (4)$$

Consequently, $\phi(y_k) \geq \phi(0)$ for all $k \geq 0$, and

$$\sum_{k=0}^{N} \|\nabla \phi(y_k)\|_*^2 \quad \leq \quad 2L(\phi)\, \mathrm{Out}(\mathcal{P}). \qquad (5)$$

Convergence of GM

Theorem. GM is well defined. For any $k \geq 0$ we have

$$\phi(y_{k+1}) - \phi(y_k) \quad \geq \quad \frac{1}{2L(\phi)} \|\nabla \phi(y_k)\|_*^2. \tag{4}$$

Consequently, $\phi(y_k) \geq \phi(0)$ for all $k \geq 0$, and

$$\sum_{k=0}^{N} \|\nabla \phi(y_k)\|_*^2 \quad \leq \quad 2L(\phi)\, \mathrm{Out}(\mathcal{P}). \tag{5}$$

Proof.

Convergence of GM

Theorem. GM is well defined. For any $k \geq 0$ we have

$$\phi(y_{k+1}) - \phi(y_k) \quad \geq \quad \tfrac{1}{2L(\phi)} \|\nabla \phi(y_k)\|_*^2. \qquad (4)$$

Consequently, $\phi(y_k) \geq \phi(0)$ for all $k \geq 0$, and

$$\sum_{k=0}^{N} \|\nabla \phi(y_k)\|_*^2 \quad \leq \quad 2L(\phi)\, \mathrm{Out}(\mathcal{P}). \qquad (5)$$

Proof. Since ϕ has Lipschitz-continuous gradient,

Convergence of GM

Theorem. GM is well defined. For any $k \geq 0$ we have

$$\phi(y_{k+1}) - \phi(y_k) \quad \geq \quad \frac{1}{2L(\phi)} \|\nabla \phi(y_k)\|_*^2. \tag{4}$$

Consequently, $\phi(y_k) \geq \phi(0)$ for all $k \geq 0$, and

$$\sum_{k=0}^{N} \|\nabla \phi(y_k)\|_*^2 \quad \leq \quad 2L(\phi)\ \mathrm{Out}(\mathcal{P}). \tag{5}$$

Proof. Since ϕ has Lipschitz-continuous gradient, we have

$$\phi(y) \quad \geq \quad \phi(y_k) + \langle \nabla \phi(y_k), y - y_k \rangle - \tfrac{1}{2} L(\phi) \|y - y_k\|^2, \quad y \in Y.$$

Convergence of GM

Theorem. GM is well defined. For any $k \geq 0$ we have

$$\phi(y_{k+1}) - \phi(y_k) \quad \geq \quad \frac{1}{2L(\phi)} \|\nabla \phi(y_k)\|_*^2. \tag{4}$$

Consequently, $\phi(y_k) \geq \phi(0)$ for all $k \geq 0$, and

$$\sum_{k=0}^{N} \|\nabla \phi(y_k)\|_*^2 \quad \leq \quad 2L(\phi)\, \mathrm{Out}(\mathcal{P}). \tag{5}$$

Proof. Since ϕ has Lipschitz-continuous gradient, we have

$$\phi(y) \quad \geq \quad \phi(y_k) + \langle \nabla \phi(y_k), y - y_k \rangle - \tfrac{1}{2} L(\phi) \|y - y_k\|^2, \quad y \in Y.$$

Substituting in this inequality $y = y'_k$,

Convergence of GM

Theorem. GM is well defined. For any $k \geq 0$ we have

$$\phi(y_{k+1}) - \phi(y_k) \quad \geq \quad \tfrac{1}{2L(\phi)} \|\nabla \phi(y_k)\|_*^2. \qquad (4)$$

Consequently, $\phi(y_k) \geq \phi(0)$ for all $k \geq 0$, and

$$\sum_{k=0}^{N} \|\nabla \phi(y_k)\|_*^2 \quad \leq \quad 2L(\phi)\, \mathrm{Out}(\mathcal{P}). \qquad (5)$$

Proof. Since ϕ has Lipschitz-continuous gradient, we have

$$\phi(y) \quad \geq \quad \phi(y_k) + \langle \nabla \phi(y_k), y - y_k \rangle - \tfrac{1}{2} L(\phi) \|y - y_k\|^2, \quad y \in Y.$$

Substituting in this inequality $y = y_k'$, we get

$$\tfrac{1}{2L(\phi)} \|\nabla \phi(y_k)\|_*^2 \leq \phi(y_k') - \phi(y_k)$$

Convergence of GM

Theorem. GM is well defined. For any $k \geq 0$ we have

$$\phi(y_{k+1}) - \phi(y_k) \;\; \geq \;\; \tfrac{1}{2L(\phi)} \|\nabla \phi(y_k)\|_*^2. \qquad (4)$$

Consequently, $\phi(y_k) \geq \phi(0)$ for all $k \geq 0$, and

$$\sum_{k=0}^{N} \|\nabla \phi(y_k)\|_*^2 \;\; \leq \;\; 2L(\phi)\, \mathrm{Out}(\mathcal{P}). \qquad (5)$$

Proof. Since ϕ has Lipschitz-continuous gradient, we have

$$\phi(y) \;\; \geq \;\; \phi(y_k) + \langle \nabla \phi(y_k), y - y_k \rangle - \tfrac{1}{2} L(\phi) \|y - y_k\|^2, \quad y \in Y.$$

Substituting in this inequality $y = y'_k$, we get

$$\tfrac{1}{2L(\phi)} \|\nabla \phi(y_k)\|_*^2 \leq \phi(y'_k) - \phi(y_k) \;\leq\; \phi(y_{k+1}) - \phi(y_k).$$

Convergence of GM

Theorem. GM is well defined. For any $k \geq 0$ we have

$$\phi(y_{k+1}) - \phi(y_k) \quad \geq \quad \tfrac{1}{2L(\phi)} \|\nabla \phi(y_k)\|_*^2. \qquad (4)$$

Consequently, $\phi(y_k) \geq \phi(0)$ for all $k \geq 0$, and

$$\sum_{k=0}^{N} \|\nabla \phi(y_k)\|_*^2 \quad \leq \quad 2L(\phi) \,\mathrm{Out}(\mathcal{P}). \qquad (5)$$

Proof. Since ϕ has Lipschitz-continuous gradient, we have

$$\phi(y) \quad \geq \quad \phi(y_k) + \langle \nabla \phi(y_k), y - y_k \rangle - \tfrac{1}{2} L(\phi) \|y - y_k\|^2, \quad y \in Y.$$

Substituting in this inequality $y = y'_k$, we get

$$\tfrac{1}{2L(\phi)} \|\nabla \phi(y_k)\|_*^2 \leq \phi(y'_k) - \phi(y_k) \;\leq\; \phi(y_{k+1}) - \phi(y_k).$$

Consider $\theta(t) = \phi(t y'_k)$.

Convergence of GM

Theorem. GM is well defined. For any $k \geq 0$ we have

$$\phi(y_{k+1}) - \phi(y_k) \quad \geq \quad \tfrac{1}{2L(\phi)} \|\nabla \phi(y_k)\|_*^2. \qquad (4)$$

Consequently, $\phi(y_k) \geq \phi(0)$ for all $k \geq 0$, and

$$\sum_{k=0}^{N} \|\nabla \phi(y_k)\|_*^2 \quad \leq \quad 2L(\phi) \, \mathrm{Out}(\mathcal{P}). \qquad (5)$$

Proof. Since ϕ has Lipschitz-continuous gradient, we have

$$\phi(y) \quad \geq \quad \phi(y_k) + \langle \nabla \phi(y_k), y - y_k \rangle - \tfrac{1}{2} L(\phi) \|y - y_k\|^2, \quad y \in Y.$$

Substituting in this inequality $y = y_k'$, we get

$$\tfrac{1}{2L(\phi)} \|\nabla \phi(y_k)\|_*^2 \leq \phi(y_k') - \phi(y_k) \; \leq \; \phi(y_{k+1}) - \phi(y_k).$$

Consider $\theta(t) = \phi(t y_k')$. Note that $\theta'(t) = \langle \nabla \phi(t y_k'), y_k' \rangle$.

Convergence of GM

Theorem. GM is well defined. For any $k \geq 0$ we have

$$\phi(y_{k+1}) - \phi(y_k) \;\geq\; \tfrac{1}{2L(\phi)} \|\nabla\phi(y_k)\|_*^2. \qquad (4)$$

Consequently, $\phi(y_k) \geq \phi(0)$ for all $k \geq 0$, and

$$\sum_{k=0}^{N} \|\nabla\phi(y_k)\|_*^2 \;\leq\; 2L(\phi)\,\mathrm{Out}(\mathcal{P}). \qquad (5)$$

Proof. Since ϕ has Lipschitz-continuous gradient, we have

$$\phi(y) \;\geq\; \phi(y_k) + \langle \nabla\phi(y_k), y - y_k \rangle - \tfrac{1}{2} L(\phi) \|y - y_k\|^2, \quad y \in Y.$$

Substituting in this inequality $y = y_k'$, we get

$$\tfrac{1}{2L(\phi)} \|\nabla\phi(y_k)\|_*^2 \leq \phi(y_k') - \phi(y_k) \;\leq\; \phi(y_{k+1}) - \phi(y_k).$$

Consider $\theta(t) = \phi(t y_k')$. Note that $\theta'(t) = \langle \nabla\phi(t y_k'), y_k' \rangle$. If $\theta'(1) \geq -\epsilon_f$, then $t_k = 1$.

Convergence of GM

Theorem. GM is well defined. For any $k \geq 0$ we have

$$\phi(y_{k+1}) - \phi(y_k) \quad \geq \quad \tfrac{1}{2L(\phi)} \|\nabla\phi(y_k)\|_*^2. \qquad (4)$$

Consequently, $\phi(y_k) \geq \phi(0)$ for all $k \geq 0$, and

$$\sum_{k=0}^{N} \|\nabla\phi(y_k)\|_*^2 \quad \leq \quad 2L(\phi)\ \mathrm{Out}(\mathcal{P}). \qquad (5)$$

Proof. Since ϕ has Lipschitz-continuous gradient, we have

$$\phi(y) \quad \geq \quad \phi(y_k) + \langle \nabla\phi(y_k), y - y_k \rangle - \tfrac{1}{2} L(\phi)\|y - y_k\|^2, \quad y \in Y.$$

Substituting in this inequality $y = y_k'$, we get

$$\tfrac{1}{2L(\phi)} \|\nabla\phi(y_k)\|_*^2 \leq \phi(y_k') - \phi(y_k) \ \leq \ \phi(y_{k+1}) - \phi(y_k).$$

Consider $\theta(t) = \phi(ty_k')$. Note that $\theta'(t) = \langle \nabla\phi(ty_k'), y_k' \rangle$. If $\theta'(1) \geq -\epsilon_f$, then $t_k = 1$. Assume $\theta'(1) < -\epsilon_f$.

Convergence of GM

Theorem. GM is well defined. For any $k \geq 0$ we have

$$\phi(y_{k+1}) - \phi(y_k) \quad \geq \quad \frac{1}{2L(\phi)} \|\nabla \phi(y_k)\|_*^2. \qquad (4)$$

Consequently, $\phi(y_k) \geq \phi(0)$ for all $k \geq 0$, and

$$\sum_{k=0}^{N} \|\nabla \phi(y_k)\|_*^2 \quad \leq \quad 2L(\phi) \, \mathrm{Out}(\mathcal{P}). \qquad (5)$$

Proof. Since ϕ has Lipschitz-continuous gradient, we have

$$\phi(y) \quad \geq \quad \phi(y_k) + \langle \nabla \phi(y_k), y - y_k \rangle - \tfrac{1}{2} L(\phi) \|y - y_k\|^2, \quad y \in Y.$$

Substituting in this inequality $y = y_k'$, we get

$$\frac{1}{2L(\phi)} \|\nabla \phi(y_k)\|_*^2 \leq \phi(y_k') - \phi(y_k) \ \leq \ \phi(y_{k+1}) - \phi(y_k).$$

Consider $\theta(t) = \phi(t y_k')$. Note that $\theta'(t) = \langle \nabla \phi(t y_k'), y_k' \rangle$. If $\theta'(1) \geq -\epsilon_f$, then $t_k = 1$. Assume $\theta'(1) < -\epsilon_f$. Note that

$\theta'(0)$

Convergence of GM

Theorem. GM is well defined. For any $k \geq 0$ we have

$$\phi(y_{k+1}) - \phi(y_k) \;\; \geq \;\; \tfrac{1}{2L(\phi)} \|\nabla \phi(y_k)\|_*^2. \qquad (4)$$

Consequently, $\phi(y_k) \geq \phi(0)$ for all $k \geq 0$, and

$$\sum_{k=0}^{N} \|\nabla \phi(y_k)\|_*^2 \;\; \leq \;\; 2L(\phi)\ \mathrm{Out}(\mathcal{P}). \qquad (5)$$

Proof. Since ϕ has Lipschitz-continuous gradient, we have

$$\phi(y) \;\; \geq \;\; \phi(y_k) + \langle \nabla \phi(y_k), y - y_k \rangle - \tfrac{1}{2} L(\phi) \|y - y_k\|^2, \quad y \in Y.$$

Substituting in this inequality $y = y_k'$, we get

$$\tfrac{1}{2L(\phi)} \|\nabla \phi(y_k)\|_*^2 \leq \phi(y_k') - \phi(y_k) \;\leq\; \phi(y_{k+1}) - \phi(y_k).$$

Consider $\theta(t) = \phi(t y_k')$. Note that $\theta'(t) = \langle \nabla \phi(t y_k'), y_k' \rangle$. If $\theta'(1) \geq -\epsilon_f$, then $t_k = 1$. Assume $\theta'(1) < -\epsilon_f$. Note that

$$\theta'(0) \geq \theta(1) - \theta(0)$$

Convergence of GM

Theorem. GM is well defined. For any $k \geq 0$ we have

$$\phi(y_{k+1}) - \phi(y_k) \quad \geq \quad \frac{1}{2L(\phi)} \|\nabla \phi(y_k)\|_*^2. \qquad (4)$$

Consequently, $\phi(y_k) \geq \phi(0)$ for all $k \geq 0$, and

$$\sum_{k=0}^{N} \|\nabla \phi(y_k)\|_*^2 \quad \leq \quad 2L(\phi)\ \mathrm{Out}(\mathcal{P}). \qquad (5)$$

Proof. Since ϕ has Lipschitz-continuous gradient, we have

$$\phi(y) \quad \geq \quad \phi(y_k) + \langle \nabla \phi(y_k), y - y_k \rangle - \tfrac{1}{2} L(\phi) \|y - y_k\|^2, \quad y \in Y.$$

Substituting in this inequality $y = y_k'$, we get

$$\frac{1}{2L(\phi)} \|\nabla \phi(y_k)\|_*^2 \leq \phi(y_k') - \phi(y_k) \ \leq \ \phi(y_{k+1}) - \phi(y_k).$$

Consider $\theta(t) = \phi(t y_k')$. Note that $\theta'(t) = \langle \nabla \phi(t y_k'), y_k' \rangle$. If $\theta'(1) \geq -\epsilon_f$, then $t_k = 1$. Assume $\theta'(1) < -\epsilon_f$. Note that

$$\theta'(0) \geq \theta(1) - \theta(0) = \phi(y_k') - \phi(0)$$

Convergence of GM

Theorem. GM is well defined. For any $k \geq 0$ we have

$$\phi(y_{k+1}) - \phi(y_k) \quad \geq \quad \frac{1}{2L(\phi)} \|\nabla \phi(y_k)\|_*^2. \qquad (4)$$

Consequently, $\phi(y_k) \geq \phi(0)$ for all $k \geq 0$, and

$$\sum_{k=0}^{N} \|\nabla \phi(y_k)\|_*^2 \quad \leq \quad 2L(\phi)\, \mathrm{Out}(\mathcal{P}). \qquad (5)$$

Proof. Since ϕ has Lipschitz-continuous gradient, we have

$$\phi(y) \quad \geq \quad \phi(y_k) + \langle \nabla \phi(y_k), y - y_k \rangle - \tfrac{1}{2} L(\phi) \|y - y_k\|^2, \quad y \in Y.$$

Substituting in this inequality $y = y_k'$, we get

$$\frac{1}{2L(\phi)} \|\nabla \phi(y_k)\|_*^2 \leq \phi(y_k') - \phi(y_k) \ \leq \ \phi(y_{k+1}) - \phi(y_k).$$

Consider $\theta(t) = \phi(t y_k')$. Note that $\theta'(t) = \langle \nabla \phi(t y_k'), y_k' \rangle$. If $\theta'(1) \geq -\epsilon_f$, then $t_k = 1$. Assume $\theta'(1) < -\epsilon_f$. Note that

$$\theta'(0) \geq \theta(1) - \theta(0) = \phi(y_k') - \phi(0) \geq \phi(y_k) - \phi(0)$$

Convergence of GM

Theorem. GM is well defined. For any $k \geq 0$ we have

$$\phi(y_{k+1}) - \phi(y_k) \quad \geq \quad \tfrac{1}{2L(\phi)} \|\nabla \phi(y_k)\|_*^2. \qquad (4)$$

Consequently, $\phi(y_k) \geq \phi(0)$ for all $k \geq 0$, and

$$\sum_{k=0}^{N} \|\nabla \phi(y_k)\|_*^2 \quad \leq \quad 2L(\phi)\ \mathrm{Out}(\mathcal{P}). \qquad (5)$$

Proof. Since ϕ has Lipschitz-continuous gradient, we have

$$\phi(y) \quad \geq \quad \phi(y_k) + \langle \nabla \phi(y_k), y - y_k \rangle - \tfrac{1}{2} L(\phi) \|y - y_k\|^2, \quad y \in Y.$$

Substituting in this inequality $y = y_k'$, we get

$$\tfrac{1}{2L(\phi)} \|\nabla \phi(y_k)\|_*^2 \leq \phi(y_k') - \phi(y_k) \ \leq \ \phi(y_{k+1}) - \phi(y_k).$$

Consider $\theta(t) = \phi(t y_k')$. Note that $\theta'(t) = \langle \nabla \phi(t y_k'), y_k' \rangle$. If $\theta'(1) \geq -\epsilon_f$, then $t_k = 1$. Assume $\theta'(1) < -\epsilon_f$. Note that

$$\theta'(0) \geq \theta(1) - \theta(0) = \phi(y_k') - \phi(0) \geq \phi(y_k) - \phi(0) \geq 0.$$

Convergence of GM

Theorem. GM is well defined. For any $k \geq 0$ we have

$$\phi(y_{k+1}) - \phi(y_k) \quad \geq \quad \frac{1}{2L(\phi)} \|\nabla \phi(y_k)\|_*^2. \tag{4}$$

Consequently, $\phi(y_k) \geq \phi(0)$ for all $k \geq 0$, and

$$\sum_{k=0}^{N} \|\nabla \phi(y_k)\|_*^2 \quad \leq \quad 2L(\phi) \, \mathrm{Out}(\mathcal{P}). \tag{5}$$

Proof. Since ϕ has Lipschitz-continuous gradient, we have

$$\phi(y) \quad \geq \quad \phi(y_k) + \langle \nabla \phi(y_k), y - y_k \rangle - \tfrac{1}{2} L(\phi) \|y - y_k\|^2, \quad y \in Y.$$

Substituting in this inequality $y = y_k'$, we get

$$\frac{1}{2L(\phi)} \|\nabla \phi(y_k)\|_*^2 \leq \phi(y_k') - \phi(y_k) \; \leq \; \phi(y_{k+1}) - \phi(y_k).$$

Consider $\theta(t) = \phi(t y_k')$. Note that $\theta'(t) = \langle \nabla \phi(t y_k'), y_k' \rangle$. If $\theta'(1) \geq -\epsilon_f$, then $t_k = 1$. Assume $\theta'(1) < -\epsilon_f$. Note that

$$\theta'(0) \geq \theta(1) - \theta(0) = \phi(y_k') - \phi(0) \geq \phi(y_k) - \phi(0) \geq 0.$$

Thus, conditions of Item 2 can be satisfied by bisection. □

Complexity of GM

Complexity of GM

Denote $k_{\#} = \arg\min_i \{ \| \nabla \phi(y_i) \|_* : 0 \le i \le k \}$.

Complexity of GM

Denote $k_{\#} = \arg\min\limits_{i} \{ \| \nabla \phi(y_i) \|_* : 0 \leq i \leq k \}$.

Corollary: $\| \nabla \phi(y_{k_{\#}}) \|_* \leq \left[\frac{2}{k+1} L(\phi) \, \mathrm{Out}(\mathcal{P}) \right]^{1/2}$, $k \geq 0$.

Complexity of GM

Denote $k_{\#} = \arg\min\limits_i \{ \|\nabla \phi(y_i)\|_* : 0 \leq i \leq k \}$.

Corollary: $\|\nabla \phi(y_{k_{\#}})\|_* \leq \left[\frac{2}{k+1} L(\phi) \, \mathrm{Out}(\mathcal{P}) \right]^{1/2}, \; k \geq 0.$

Hence, for satisfying conditions (3),

Complexity of GM

Denote $k_{\#} = \arg\min_i \{ \| \nabla \phi(y_i) \|_* : 0 \leq i \leq k \}$.

Corollary: $\| \nabla \phi(y_{k_{\#}}) \|_* \leq \left[\frac{2}{k+1} L(\phi) \, \mathrm{Out}(\mathcal{P}) \right]^{1/2}, \; k \geq 0.$

Hence, for satisfying conditions (3), we need at most

$$N_{\epsilon_g} \stackrel{\mathrm{def}}{=} \frac{2}{\epsilon_g^2} L(\phi) \, \mathrm{Out}(\mathcal{P}) \quad \text{iterations.} \tag{6}$$

Complexity of GM

Denote $k_\# = \arg\min_i \{\|\nabla\phi(y_i)\|_* : 0 \le i \le k\}$.

Corollary: $\|\nabla\phi(y_{k_\#})\|_* \le \left[\frac{2}{k+1} L(\phi)\,\mathrm{Out}(\mathcal{P})\right]^{1/2}$, $k \ge 0$.

Hence, for satisfying conditions (3), we need at most

$$N_{\epsilon_g} \stackrel{\mathrm{def}}{=} \frac{2}{\epsilon_g^2} L(\phi)\,\mathrm{Out}(\mathcal{P}) \quad \text{iterations.} \tag{6}$$

Complexity of 1D-max. of $\phi(ty_k')$ depends on the size of y_k'.

Complexity of GM

Denote $k_{\#} = \arg\min_i \{\|\nabla\phi(y_i)\|_* : 0 \le i \le k\}$.

Corollary: $\|\nabla\phi(y_{k_{\#}})\|_* \le \left[\frac{2}{k+1} L(\phi)\,\mathrm{Out}(\mathcal{P})\right]^{1/2}, \; k \ge 0.$

Hence, for satisfying conditions (3), we need at most

$$N_{\epsilon_g} \stackrel{\mathrm{def}}{=} \frac{2}{\epsilon_g^2} L(\phi)\,\mathrm{Out}(\mathcal{P}) \quad \text{iterations.} \tag{6}$$

Complexity of 1D-max. of $\phi(ty_k')$ depends on the size of y_k'.

Lemma. $\|y_{k+1}\| \le \|y_k'\| \le \left[\frac{k+1}{2L(\phi)}\mathrm{Out}(\mathcal{P})\right]^{1/2}$.

Complexity of GM

Denote $k_{\#} = \arg\min\limits_i \{ \|\nabla \phi(y_i)\|_* : 0 \leq i \leq k \}$.

Corollary: $\|\nabla \phi(y_{k_{\#}})\|_* \leq \left[\frac{2}{k+1} L(\phi) \, \mathrm{Out}(\mathcal{P}) \right]^{1/2}$, $k \geq 0$.

Hence, for satisfying conditions (3), we need at most

$$N_{\epsilon_g} \stackrel{\mathrm{def}}{=} \frac{2}{\epsilon_g^2} L(\phi) \, \mathrm{Out}(\mathcal{P}) \quad \text{iterations.} \tag{6}$$

Complexity of 1D-max. of $\phi(t y'_k)$ depends on the size of y'_k.

Lemma. $\|y_{k+1}\| \leq \|y'_k\| \leq \left[\frac{k+1}{2L(\phi)} \mathrm{Out}(\mathcal{P}) \right]^{1/2}$.

Proof.

Complexity of GM

Denote $k_{\#} = \arg\min_i \{\|\nabla\phi(y_i)\|_* : 0 \le i \le k\}$.

Corollary: $\|\nabla\phi(y_{k_{\#}})\|_* \le \left[\frac{2}{k+1} L(\phi)\,\mathrm{Out}(\mathcal{P})\right]^{1/2}, \; k \ge 0.$

Hence, for satisfying conditions (3), we need at most

$$N_{\epsilon_g} \stackrel{\mathrm{def}}{=} \frac{2}{\epsilon_g^2} L(\phi)\,\mathrm{Out}(\mathcal{P}) \quad \text{iterations.} \tag{6}$$

Complexity of 1D-max. of $\phi(ty_k')$ depends on the size of y_k'.

Lemma. $\|y_{k+1}\| \le \|y_k'\| \le \left[\frac{k+1}{2L(\phi)}\mathrm{Out}(\mathcal{P})\right]^{1/2}.$

Proof. Since $t_k \le 1$,

Complexity of GM

Denote $k_\# = \arg\min_i \{ \|\nabla \phi(y_i)\|_* : 0 \leq i \leq k \}$.

Corollary: $\|\nabla \phi(y_{k_\#})\|_* \leq \left[\frac{2}{k+1} L(\phi) \, \mathrm{Out}(\mathcal{P}) \right]^{1/2}, \; k \geq 0.$

Hence, for satisfying conditions (3), we need at most

$$N_{\epsilon_g} \stackrel{\mathrm{def}}{=} \frac{2}{\epsilon_g^2} L(\phi) \, \mathrm{Out}(\mathcal{P}) \quad \text{iterations.} \tag{6}$$

Complexity of 1D-max. of $\phi(t y_k')$ depends on the size of y_k'.

Lemma. $\|y_{k+1}\| \leq \|y_k'\| \leq \left[\frac{k+1}{2L(\phi)} \mathrm{Out}(\mathcal{P}) \right]^{1/2}$.

Proof. Since $t_k \leq 1$,

$\|y_{k+1}\|$

Complexity of GM

Denote $k_{\#} = \arg\min_i \{\|\nabla\phi(y_i)\|_* : 0 \leq i \leq k\}$.

Corollary: $\|\nabla\phi(y_{k_{\#}})\|_* \leq \left[\frac{2}{k+1} L(\phi)\, \mathrm{Out}(\mathcal{P})\right]^{1/2}$, $k \geq 0$.

Hence, for satisfying conditions (3), we need at most

$$N_{\epsilon_g} \stackrel{\mathrm{def}}{=} \frac{2}{\epsilon_g^2} L(\phi)\, \mathrm{Out}(\mathcal{P}) \quad \text{iterations.} \tag{6}$$

Complexity of 1D-max. of $\phi(t y_k')$ depends on the size of y_k'.

Lemma. $\|y_{k+1}\| \leq \|y_k'\| \leq \left[\frac{k+1}{2L(\phi)} \mathrm{Out}(\mathcal{P})\right]^{1/2}$.

Proof. Since $t_k \leq 1$,

$\|y_{k+1}\| \leq \|y_k'\|$

Complexity of GM

Denote $k_{\#} = \arg\min_i \{\|\nabla\phi(y_i)\|_* : 0 \leq i \leq k\}$.

Corollary: $\|\nabla\phi(y_{k_{\#}})\|_* \leq \left[\frac{2}{k+1} L(\phi)\,\mathrm{Out}(\mathcal{P})\right]^{1/2}$, $k \geq 0$.

Hence, for satisfying conditions (3), we need at most

$$N_{\epsilon_g} \stackrel{\mathrm{def}}{=} \frac{2}{\epsilon_g^2} L(\phi)\,\mathrm{Out}(\mathcal{P}) \quad \text{iterations.} \tag{6}$$

Complexity of 1D-max. of $\phi(ty'_k)$ depends on the size of y'_k.

Lemma. $\|y_{k+1}\| \leq \|y'_k\| \leq \left[\frac{k+1}{2L(\phi)}\mathrm{Out}(\mathcal{P})\right]^{1/2}$.

Proof. Since $t_k \leq 1$,

$\|y_{k+1}\| \leq \|y'_k\| \leq \|y_k\| + \frac{1}{2L(\phi)}\|\nabla\phi(y_k)\|_*$.

Complexity of GM

Denote $k_{\#} = \arg\min_i \{\|\nabla \phi(y_i)\|_* : 0 \leq i \leq k\}$.

Corollary: $\|\nabla \phi(y_{k_{\#}})\|_* \leq \left[\frac{2}{k+1} L(\phi) \, \mathrm{Out}(\mathcal{P}) \right]^{1/2}$, $k \geq 0$.

Hence, for satisfying conditions (3), we need at most

$$N_{\epsilon_g} \stackrel{\mathrm{def}}{=} \frac{2}{\epsilon_g^2} L(\phi) \, \mathrm{Out}(\mathcal{P}) \quad \text{iterations.} \qquad (6)$$

Complexity of 1D-max. of $\phi(t y_k')$ depends on the size of y_k'.

Lemma. $\|y_{k+1}\| \leq \|y_k'\| \leq \left[\frac{k+1}{2L(\phi)} \mathrm{Out}(\mathcal{P}) \right]^{1/2}$.

Proof. Since $t_k \leq 1$,

$\|y_{k+1}\| \leq \|y_k'\| \leq \|y_k\| + \frac{1}{2L(\phi)} \|\nabla \phi(y_k)\|_*$.

Hence, $\|y_{k+1}\| \leq \frac{1}{2L(\phi)} \sum_{i=0}^{k} \|\nabla \phi(y_k)\|_*$.

Complexity of GM

Denote $k_{\#} = \arg\min_i \{\|\nabla\phi(y_i)\|_* : 0 \le i \le k\}$.

Corollary: $\|\nabla\phi(y_{k_{\#}})\|_* \le \left[\frac{2}{k+1} L(\phi)\,\mathrm{Out}(\mathcal{P})\right]^{1/2}$, $k \ge 0$.

Hence, for satisfying conditions (3), we need at most

$$N_{\epsilon_g} \stackrel{\mathrm{def}}{=} \frac{2}{\epsilon_g^2} L(\phi)\,\mathrm{Out}(\mathcal{P}) \quad \text{iterations.} \tag{6}$$

Complexity of 1D-max. of $\phi(t y_k')$ depends on the size of y_k'.

Lemma. $\|y_{k+1}\| \le \|y_k'\| \le \left[\frac{k+1}{2L(\phi)}\mathrm{Out}(\mathcal{P})\right]^{1/2}$.

Proof. Since $t_k \le 1$,

$\|y_{k+1}\| \le \|y_k'\| \le \|y_k\| + \frac{1}{2L(\phi)}\|\nabla\phi(y_k)\|_*$.

Hence, $\|y_{k+1}\| \le \frac{1}{2L(\phi)} \sum\limits_{i=0}^{k} \|\nabla\phi(y_k)\|_*$. It remains to use (5). □

Complexity of GM

Denote $k_{\#} = \arg\min_i \{ \|\nabla \phi(y_i)\|_* : 0 \leq i \leq k \}$.

Corollary: $\|\nabla \phi(y_{k_{\#}})\|_* \leq \left[\frac{2}{k+1} L(\phi) \, \mathrm{Out}(\mathcal{P}) \right]^{1/2}$, $k \geq 0$.

Hence, for satisfying conditions (3), we need at most

$$N_{\epsilon_g} \overset{\mathrm{def}}{=} \frac{2}{\epsilon_g^2} L(\phi) \, \mathrm{Out}(\mathcal{P}) \quad \text{iterations.} \tag{6}$$

Complexity of 1D-max. of $\phi(ty'_k)$ depends on the size of y'_k.

Lemma. $\|y_{k+1}\| \leq \|y'_k\| \leq \left[\frac{k+1}{2L(\phi)} \mathrm{Out}(\mathcal{P}) \right]^{1/2}$.

Proof. Since $t_k \leq 1$,

$\|y_{k+1}\| \leq \|y'_k\| \ \leq \ \|y_k\| + \frac{1}{2L(\phi)} \|\nabla \phi(y_k)\|_*$.

Hence, $\|y_{k+1}\| \leq \frac{1}{2L(\phi)} \sum\limits_{i=0}^{k} \|\nabla \phi(y_k)\|_*$. It remains to use (5). $\square$

Substituting in this bound the iteration bound (6),

Complexity of GM

Denote $k_{\#} = \arg\min_i \{ \| \nabla \phi(y_i) \|_* : 0 \le i \le k \}$.

Corollary: $\| \nabla \phi(y_{k_{\#}}) \|_* \le \left[\frac{2}{k+1} L(\phi) \, \mathrm{Out}(\mathcal{P}) \right]^{1/2}$, $k \ge 0$.

Hence, for satisfying conditions (3), we need at most

$$N_{\epsilon_g} \stackrel{\mathrm{def}}{=} \frac{2}{\epsilon_g^2} L(\phi) \, \mathrm{Out}(\mathcal{P}) \quad \text{iterations.} \qquad (6)$$

Complexity of 1D-max. of $\phi(t y_k')$ depends on the size of y_k'.

Lemma. $\| y_{k+1} \| \le \| y_k' \| \le \left[\frac{k+1}{2L(\phi)} \mathrm{Out}(\mathcal{P}) \right]^{1/2}$.

Proof. Since $t_k \le 1$,

$\| y_{k+1} \| \le \| y_k' \| \le \| y_k \| + \frac{1}{2L(\phi)} \| \nabla \phi(y_k) \|_*$.

Hence, $\| y_{k+1} \| \le \frac{1}{2L(\phi)} \sum_{i=0}^{k} \| \nabla \phi(y_k) \|_*$. It remains to use (5). □

Substituting in this bound the iteration bound (6), we obtain

$$\| y_{k+1} \| \le \frac{1}{\epsilon_g} \mathrm{Out}(\mathcal{P}), \quad 0 \le k \le N_{\epsilon_g}.$$

Fast Gradient Method

Fast Gradient Method

Let us choose $B \succ 0$.

Fast Gradient Method

Let us choose $B \succ 0$. Define $\|y\| = \langle By, y \rangle^{1/2}$.

Fast Gradient Method

Let us choose $B \succ 0$. Define $\|y\| = \langle By, y \rangle^{1/2}$.
For a fixed $\delta > 0$, denote $\phi_\delta(y) = \phi(y) - \frac{\delta}{2}\|y\|^2$.

Fast Gradient Method

Let us choose $B \succ 0$. Define $\|y\| = \langle By, y\rangle^{1/2}$.
For a fixed $\delta > 0$, denote $\phi_\delta(y) = \phi(y) - \frac{\delta}{2}\|y\|^2$.

Problem: $\phi_\delta^* \stackrel{\text{def}}{=} \max\limits_{y \in Y} \phi_\delta(y)$.

Fast Gradient Method

Let us choose $B \succ 0$. Define $\|y\| = \langle By, y \rangle^{1/2}$.
For a fixed $\delta > 0$, denote $\phi_\delta(y) = \phi(y) - \frac{\delta}{2}\|y\|^2$.

Problem: $\phi_\delta^* \stackrel{\text{def}}{=} \max\limits_{y \in Y} \phi_\delta(y)$.

Denote by y_δ^* its unique optimal solution.

Fast Gradient Method

Let us choose $B \succ 0$. Define $\|y\| = \langle By, y\rangle^{1/2}$.
For a fixed $\delta > 0$, denote $\phi_\delta(y) = \phi(y) - \frac{\delta}{2}\|y\|^2$.

Problem: $\phi_\delta^* \stackrel{\mathrm{def}}{=} \max\limits_{y \in Y} \phi_\delta(y)$.

Denote by y_δ^* its unique optimal solution. Note that

$$\phi_\delta^*$$

Fast Gradient Method

Let us choose $B \succ 0$. Define $\|y\| = \langle By, y \rangle^{1/2}$.
For a fixed $\delta > 0$, denote $\phi_\delta(y) = \phi(y) - \frac{\delta}{2}\|y\|^2$.

Problem: $\phi_\delta^* \stackrel{\mathrm{def}}{=} \max\limits_{y \in Y} \phi_\delta(y)$.

Denote by y_δ^* its unique optimal solution. Note that

$$\phi_\delta^* = \phi(y_\delta^*) - \tfrac{\delta}{2}\|y_\delta^*\|^2$$

Fast Gradient Method

Let us choose $B \succ 0$. Define $\|y\| = \langle By, y \rangle^{1/2}$.
For a fixed $\delta > 0$, denote $\phi_\delta(y) = \phi(y) - \frac{\delta}{2}\|y\|^2$.

Problem: $\phi_\delta^* \stackrel{\mathrm{def}}{=} \max\limits_{y \in Y} \phi_\delta(y)$.

Denote by y_δ^* its unique optimal solution. Note that

$$\phi_\delta^* = \phi(y_\delta^*) - \frac{\delta}{2}\|y_\delta^*\|^2 \leq \phi^* - \frac{\delta}{2}\|y_\delta^*\|^2.$$

Fast Gradient Method

Let us choose $B \succ 0$. Define $\|y\| = \langle By, y \rangle^{1/2}$.
For a fixed $\delta > 0$, denote $\phi_\delta(y) = \phi(y) - \frac{\delta}{2}\|y\|^2$.

Problem: $\phi_\delta^* \stackrel{\mathrm{def}}{=} \max\limits_{y \in Y} \phi_\delta(y)$.

Denote by y_δ^* its unique optimal solution. Note that

$$\phi_\delta^* = \phi(y_\delta^*) - \frac{\delta}{2}\|y_\delta^*\|^2 \leq \phi^* - \frac{\delta}{2}\|y_\delta^*\|^2.$$

Therefore, $\frac{\delta}{2}\|y_\delta^*\|^2$

Fast Gradient Method

Let us choose $B \succ 0$. Define $\|y\| = \langle By, y\rangle^{1/2}$.
For a fixed $\delta > 0$, denote $\phi_\delta(y) = \phi(y) - \frac{\delta}{2}\|y\|^2$.

Problem: $\phi_\delta^* \stackrel{\mathrm{def}}{=} \max\limits_{y \in Y} \phi_\delta(y)$.

Denote by y_δ^* its unique optimal solution. Note that

$$\phi_\delta^* = \phi(y_\delta^*) - \frac{\delta}{2}\|y_\delta^*\|^2 \leq \phi^* - \frac{\delta}{2}\|y_\delta^*\|^2.$$

Therefore, $\frac{\delta}{2}\|y_\delta^*\|^2 \leq \phi_\delta^* - \phi_\delta(0)$

Fast Gradient Method

Let us choose $B \succ 0$. Define $\|y\| = \langle By, y\rangle^{1/2}$.
For a fixed $\delta > 0$, denote $\phi_\delta(y) = \phi(y) - \frac{\delta}{2}\|y\|^2$.

Problem: $\phi^*_\delta \stackrel{\mathrm{def}}{=} \max\limits_{y \in Y} \phi_\delta(y)$.

Denote by y^*_δ its unique optimal solution. Note that

$$\phi^*_\delta = \phi(y^*_\delta) - \frac{\delta}{2}\|y^*_\delta\|^2 \leq \phi^* - \frac{\delta}{2}\|y^*_\delta\|^2.$$

Therefore, $\frac{\delta}{2}\|y^*_\delta\|^2 \leq \phi^*_\delta - \phi_\delta(0) \stackrel{(2)}{\leq} \mathrm{Out}(\mathcal{P}) - \frac{\delta}{2}\|y^*_\delta\|^2$.

Fast Gradient Method

Let us choose $B \succ 0$. Define $\|y\| = \langle By, y\rangle^{1/2}$.
For a fixed $\delta > 0$, denote $\phi_\delta(y) = \phi(y) - \frac{\delta}{2}\|y\|^2$.

Problem: $\phi_\delta^* \stackrel{\mathrm{def}}{=} \max\limits_{y \in Y} \phi_\delta(y)$.

Denote by y_δ^* its unique optimal solution. Note that

$$\phi_\delta^* = \phi(y_\delta^*) - \frac{\delta}{2}\|y_\delta^*\|^2 \le \phi^* - \frac{\delta}{2}\|y_\delta^*\|^2.$$

Therefore, $\frac{\delta}{2}\|y_\delta^*\|^2 \le \phi_\delta^* - \phi_\delta(0) \stackrel{(2)}{\le} \mathrm{Out}(\mathcal{P}) - \frac{\delta}{2}\|y_\delta^*\|^2$.

NB: $L(\phi_\delta) = L(\phi) + \delta$, and $\sigma(\phi_\delta) = \delta$.

Fast Gradient Method

Let us choose $B \succ 0$. Define $\|y\| = \langle By, y\rangle^{1/2}$.
For a fixed $\delta > 0$, denote $\phi_\delta(y) = \phi(y) - \frac{\delta}{2}\|y\|^2$.

Problem: $\phi_\delta^* \stackrel{\mathrm{def}}{=} \max\limits_{y \in Y} \phi_\delta(y)$.

Denote by y_δ^* its unique optimal solution. Note that

$$\phi_\delta^* = \phi(y_\delta^*) - \frac{\delta}{2}\|y_\delta^*\|^2 \leq \phi^* - \frac{\delta}{2}\|y_\delta^*\|^2.$$

Therefore, $\frac{\delta}{2}\|y_\delta^*\|^2 \leq \phi_\delta^* - \phi_\delta(0) \stackrel{(2)}{\leq} \mathrm{Out}(\mathcal{P}) - \frac{\delta}{2}\|y_\delta^*\|^2$.

NB: $L(\phi_\delta) = L(\phi) + \delta$, and $\sigma(\phi_\delta) = \delta$.

Let us apply the simplest FGM with starting point $y_0 = u_0 = 0$:

Fast Gradient Method

Let us choose $B \succ 0$. Define $\|y\| = \langle By, y\rangle^{1/2}$.
For a fixed $\delta > 0$, denote $\phi_\delta(y) = \phi(y) - \frac{\delta}{2}\|y\|^2$.

Problem: $\phi_\delta^* \stackrel{\mathrm{def}}{=} \max\limits_{y \in Y} \phi_\delta(y)$.

Denote by y_δ^* its unique optimal solution. Note that

$$\phi_\delta^* = \phi(y_\delta^*) - \frac{\delta}{2}\|y_\delta^*\|^2 \le \phi^* - \frac{\delta}{2}\|y_\delta^*\|^2.$$

Therefore, $\frac{\delta}{2}\|y_\delta^*\|^2 \le \phi_\delta^* - \phi_\delta(0) \stackrel{(2)}{\le} \mathrm{Out}(\mathcal{P}) - \frac{\delta}{2}\|y_\delta^*\|^2$.

NB: $L(\phi_\delta) = L(\phi) + \delta$, and $\sigma(\phi_\delta) = \delta$.

Let us apply the simplest FGM with starting point $y_0 = u_0 = 0$:

$$y_{k+1} = u_k + \frac{1}{L(\phi)+\delta} B^{-1} \nabla \phi_\delta(u_k),$$

Fast Gradient Method

Let us choose $B \succ 0$. Define $\|y\| = \langle By, y\rangle^{1/2}$.
For a fixed $\delta > 0$, denote $\phi_\delta(y) = \phi(y) - \frac{\delta}{2}\|y\|^2$.

Problem: $\phi_\delta^* \stackrel{\mathrm{def}}{=} \max\limits_{y \in Y} \phi_\delta(y)$.

Denote by y_δ^* its unique optimal solution. Note that

$$\phi_\delta^* = \phi(y_\delta^*) - \tfrac{\delta}{2}\|y_\delta^*\|^2 \leq \phi^* - \tfrac{\delta}{2}\|y_\delta^*\|^2.$$

Therefore, $\frac{\delta}{2}\|y_\delta^*\|^2 \leq \phi_\delta^* - \phi_\delta(0) \stackrel{(2)}{\leq} \mathrm{Out}(\mathcal{P}) - \frac{\delta}{2}\|y_\delta^*\|^2$.

NB: $L(\phi_\delta) = L(\phi) + \delta$, and $\sigma(\phi_\delta) = \delta$.

Let us apply the simplest FGM with starting point $y_0 = u_0 = 0$:

$$y_{k+1} = u_k + \tfrac{1}{L(\phi)+\delta} B^{-1}\nabla\phi_\delta(u_k),$$

$$u_{k+1} = y_{k+1} + \kappa(y_{k+1} - y_k),$$

Fast Gradient Method

Let us choose $B \succ 0$. Define $\|y\| = \langle By, y\rangle^{1/2}$.
For a fixed $\delta > 0$, denote $\phi_\delta(y) = \phi(y) - \frac{\delta}{2}\|y\|^2$.

Problem: $\phi_\delta^* \stackrel{\text{def}}{=} \max\limits_{y \in Y} \phi_\delta(y)$.

Denote by y_δ^* its unique optimal solution. Note that

$$\phi_\delta^* = \phi(y_\delta^*) - \tfrac{\delta}{2}\|y_\delta^*\|^2 \leq \phi^* - \tfrac{\delta}{2}\|y_\delta^*\|^2.$$

Therefore, $\frac{\delta}{2}\|y_\delta^*\|^2 \leq \phi_\delta^* - \phi_\delta(0) \stackrel{(2)}{\leq} \text{Out}(\mathcal{P}) - \frac{\delta}{2}\|y_\delta^*\|^2$.

NB: $L(\phi_\delta) = L(\phi) + \delta$, and $\sigma(\phi_\delta) = \delta$.

Let us apply the simplest FGM with starting point $y_0 = u_0 = 0$:

$$y_{k+1} = u_k + \tfrac{1}{L(\phi)+\delta} B^{-1}\nabla\phi_\delta(u_k),$$

$$u_{k+1} = y_{k+1} + \kappa(y_{k+1} - y_k),$$

where $\kappa = \frac{[L(\phi)+\delta]^{1/2} - \delta^{1/2}}{[L(\phi)+\delta]^{1/2} + \delta^{1/2}}$.

Convergence of FGM

Convergence of FGM

$$\phi_\delta^* - \phi_\delta(y_k) \quad \leq \quad 2\exp\left(-k\sqrt{\tfrac{\delta}{L(\phi)+\delta}}\right)\cdot \mathrm{Out}(\mathcal{P}).$$

Convergence of FGM

$$\phi_\delta^* - \phi_\delta(y_k) \quad \leq \quad 2\exp\left(-k\sqrt{\tfrac{\delta}{L(\phi)+\delta}}\right)\cdot \mathrm{Out}(\mathcal{P}).$$

Thus, for any $\xi > 0$ and k large enough, we can ensure

$$\xi \quad \geq \quad \phi_\delta^* - \phi_\delta(y_k)$$

Convergence of FGM

$$\phi_\delta^* - \phi_\delta(y_k) \quad \leq \quad 2\exp\left(-k\sqrt{\frac{\delta}{L(\phi)+\delta}}\right) \cdot \mathrm{Out}(\mathcal{P}).$$

Thus, for any $\xi > 0$ and k large enough, we can ensure

$$\xi \quad \geq \quad \phi_\delta^* - \phi_\delta(y_k) \;\geq\; \frac{1}{2(L(\phi)+\delta)}\|\nabla\phi_\delta(y_k)\|_*^2$$

Convergence of FGM

$$\phi_\delta^* - \phi_\delta(y_k) \quad \leq \quad 2\exp\left(-k\sqrt{\tfrac{\delta}{L(\phi)+\delta}}\right)\cdot \mathrm{Out}(\mathcal{P}).$$

Thus, for any $\xi > 0$ and k large enough, we can ensure

$$\begin{aligned} \xi \quad &\geq \quad \phi_\delta^* - \phi_\delta(y_k) \;\geq\; \tfrac{1}{2(L(\phi)+\delta)}\|\nabla\phi_\delta(y_k)\|_*^2 \\ &= \quad \tfrac{1}{2(L(\phi)+\delta)}\|\nabla\phi(y_k) - \delta B y_k\|_*^2 \end{aligned}$$

Convergence of FGM

$$\phi_\delta^* - \phi_\delta(y_k) \quad \leq \quad 2\exp\left(-k\sqrt{\tfrac{\delta}{L(\phi)+\delta}}\right) \cdot \mathrm{Out}(\mathcal{P}).$$

Thus, for any $\xi > 0$ and k large enough, we can ensure

$$\begin{aligned}
\xi \quad &\geq \quad \phi_\delta^* - \phi_\delta(y_k) \;\geq\; \tfrac{1}{2(L(\phi)+\delta)} \|\nabla \phi_\delta(y_k)\|_*^2 \\
&= \quad \tfrac{1}{2(L(\phi)+\delta)} \|\nabla \phi(y_k) - \delta B y_k\|_*^2 \;\geq\; -\tfrac{2\delta}{L(\phi)+\delta} \langle \nabla \phi(y_k), y_k \rangle.
\end{aligned}$$

Convergence of FGM

$$\phi_\delta^* - \phi_\delta(y_k) \quad \leq \quad 2\exp\left(-k\sqrt{\tfrac{\delta}{L(\phi)+\delta}}\right) \cdot \mathrm{Out}(\mathcal{P}).$$

Thus, for any $\xi > 0$ and k large enough, we can ensure

$$\begin{array}{rcl} \xi & \geq & \phi_\delta^* - \phi_\delta(y_k) \ \geq \ \frac{1}{2(L(\phi)+\delta)}\|\nabla\phi_\delta(y_k)\|_*^2 \\ & = & \frac{1}{2(L(\phi)+\delta)}\|\nabla\phi(y_k) - \delta B y_k\|_*^2 \ \geq \ -\frac{2\delta}{L(\phi)+\delta}\langle\nabla\phi(y_k), y_k\rangle. \end{array}$$

Thus, in order to guarantee $(3)_b$, we need $\xi \leq \frac{2\,\epsilon_f\,\delta}{L(\phi)+\delta}$.

Convergence of FGM

$$\phi_\delta^* - \phi_\delta(y_k) \quad \leq \quad 2\exp\left(-k\sqrt{\tfrac{\delta}{L(\phi)+\delta}}\right) \cdot \mathrm{Out}(\mathcal{P}).$$

Thus, for any $\xi > 0$ and k large enough, we can ensure

$$\begin{aligned} \xi \quad &\geq \quad \phi_\delta^* - \phi_\delta(y_k) \;\geq\; \tfrac{1}{2(L(\phi)+\delta)}\|\nabla\phi_\delta(y_k)\|_*^2 \\ &= \quad \tfrac{1}{2(L(\phi)+\delta)}\|\nabla\phi(y_k) - \delta B y_k\|_*^2 \;\geq\; -\tfrac{2\delta}{L(\phi)+\delta}\langle\nabla\phi(y_k), y_k\rangle. \end{aligned}$$

Thus, in order to guarantee $(3)_b$, we need $\xi \leq \frac{2\,\epsilon_f\,\delta}{L(\phi)+\delta}$.
On the other hand, for $(3)_a$ we need

$$\|\nabla\phi(y_k)\|_*$$

Convergence of FGM

$$\phi_\delta^* - \phi_\delta(y_k) \quad \leq \quad 2\exp\left(-k\sqrt{\tfrac{\delta}{L(\phi)+\delta}}\right) \cdot \mathrm{Out}(\mathcal{P}).$$

Thus, for any $\xi > 0$ and k large enough, we can ensure

$$\begin{aligned} \xi \quad &\geq \quad \phi_\delta^* - \phi_\delta(y_k) \;\geq\; \tfrac{1}{2(L(\phi)+\delta)}\|\nabla\phi_\delta(y_k)\|_*^2 \\ &= \quad \tfrac{1}{2(L(\phi)+\delta)}\|\nabla\phi(y_k) - \delta B y_k\|_*^2 \;\geq\; -\tfrac{2\delta}{L(\phi)+\delta}\langle\nabla\phi(y_k), y_k\rangle. \end{aligned}$$

Thus, in order to guarantee $(3)_b$, we need $\xi \leq \frac{2\,\epsilon_f\,\delta}{L(\phi)+\delta}$.
On the other hand, for $(3)_a$ we need

$$\|\nabla\phi(y_k)\|_* \quad \leq \quad [2\xi(L(\phi)+\delta)]^{1/2} + \delta\|y_k\|$$

Convergence of FGM

$$\phi_\delta^* - \phi_\delta(y_k) \quad \leq \quad 2\exp\left(-k\sqrt{\tfrac{\delta}{L(\phi)+\delta}}\right) \cdot \mathrm{Out}(\mathcal{P}).$$

Thus, for any $\xi > 0$ and k large enough, we can ensure

$$\begin{array}{rcl} \xi & \geq & \phi_\delta^* - \phi_\delta(y_k) \;\geq\; \frac{1}{2(L(\phi)+\delta)} \|\nabla \phi_\delta(y_k)\|_*^2 \\ & = & \frac{1}{2(L(\phi)+\delta)} \|\nabla \phi(y_k) - \delta B y_k\|_*^2 \;\geq\; -\frac{2\delta}{L(\phi)+\delta} \langle \nabla \phi(y_k), y_k \rangle. \end{array}$$

Thus, in order to guarantee $(3)_b$, we need $\xi \leq \frac{2\,\epsilon_f\,\delta}{L(\phi)+\delta}$.
On the other hand, for $(3)_a$ we need

$$\begin{array}{rcl} \|\nabla \phi(y_k)\|_* & \leq & [2\xi(L(\phi)+\delta)]^{1/2} + \delta\|y_k\| \\ & \leq & [2\xi(L(\phi)+\delta)]^{1/2} + \delta\left[\frac{1}{\delta}\mathrm{Out}(\mathcal{P})\right]^{1/2} \end{array}$$

Convergence of FGM

$$\phi_\delta^* - \phi_\delta(y_k) \;\; \leq \;\; 2\exp\left(-k\sqrt{\tfrac{\delta}{L(\phi)+\delta}}\right) \cdot \mathrm{Out}(\mathcal{P}).$$

Thus, for any $\xi > 0$ and k large enough, we can ensure

$$\begin{array}{rcl} \xi & \geq & \phi_\delta^* - \phi_\delta(y_k) \;\geq\; \frac{1}{2(L(\phi)+\delta)}\|\nabla\phi_\delta(y_k)\|_*^2 \\ & = & \frac{1}{2(L(\phi)+\delta)}\|\nabla\phi(y_k) - \delta B y_k\|_*^2 \;\geq\; -\frac{2\delta}{L(\phi)+\delta}\langle\nabla\phi(y_k), y_k\rangle. \end{array}$$

Thus, in order to guarantee $(3)_b$, we need $\xi \leq \frac{2\,\epsilon_f\,\delta}{L(\phi)+\delta}$.
On the other hand, for $(3)_a$ we need

$$\begin{array}{rcl} \|\nabla\phi(y_k)\|_* & \leq & [2\xi(L(\phi)+\delta)]^{1/2} + \delta\|y_k\| \\ & \leq & [2\xi(L(\phi)+\delta)]^{1/2} + \delta\left[\frac{1}{\delta}\mathrm{Out}(\mathcal{P})\right]^{1/2} \;\leq\; \epsilon_g. \end{array}$$

Convergence of FGM

$$\phi_\delta^* - \phi_\delta(y_k) \;\; \leq \;\; 2\exp\left(-k\sqrt{\tfrac{\delta}{L(\phi)+\delta}}\right) \cdot \mathrm{Out}(\mathcal{P}).$$

Thus, for any $\xi > 0$ and k large enough, we can ensure

$$\begin{array}{rcl} \xi & \geq & \phi_\delta^* - \phi_\delta(y_k) \;\geq\; \frac{1}{2(L(\phi)+\delta)}\|\nabla \phi_\delta(y_k)\|_*^2 \\ & = & \frac{1}{2(L(\phi)+\delta)}\|\nabla \phi(y_k) - \delta B y_k\|_*^2 \;\geq\; -\frac{2\delta}{L(\phi)+\delta}\langle \nabla\phi(y_k), y_k\rangle. \end{array}$$

Thus, in order to guarantee $(3)_b$, we need $\xi \leq \frac{2\,\epsilon_f\,\delta}{L(\phi)+\delta}$.
On the other hand, for $(3)_a$ we need

$$\begin{array}{rcl} \|\nabla\phi(y_k)\|_* & \leq & [2\xi(L(\phi)+\delta)]^{1/2} + \delta\|y_k\| \\ & \leq & [2\xi(L(\phi)+\delta)]^{1/2} + \delta\left[\frac{1}{\delta}\mathrm{Out}(\mathcal{P})\right]^{1/2} \;\leq\; \epsilon_g. \end{array}$$

Hence, ξ must satisfy inequality

$$\xi \;\; \leq \;\; \min\{\tfrac{2\,\epsilon_f\,\delta}{L(\phi)+\delta}, \tfrac{(\epsilon_g - [\delta \mathrm{Out}(\mathcal{P})]^{1/2})^2}{2(L(\phi)+\delta)}\}.$$

Complexity of FGM

Complexity of FGM

Choice of ξ:

Complexity of FGM

Choice of ξ: for $\sqrt{\delta} = \frac{\epsilon_g}{\text{Out}^{1/2} + 2\epsilon_f^{1/2}}$

Complexity of FGM

Choice of ξ: for $\sqrt{\delta} = \frac{\epsilon_g}{\mathrm{Out}^{1/2} + 2\epsilon_f^{1/2}} \approx \frac{\epsilon_g}{\mathrm{Out}^{1/2}(\mathcal{P})}$,

Complexity of FGM

Choice of ξ: for $\sqrt{\delta} = \frac{\epsilon_g}{\mathrm{Out}^{1/2} + 2\epsilon_f^{1/2}} \approx \frac{\epsilon_g}{\mathrm{Out}^{1/2}(\mathcal{P})}$, we get

$$\xi \approx \frac{2\,\epsilon_f\,\epsilon_g^2}{L(\phi)\mathrm{Out}(\mathcal{P})}.$$

Complexity of FGM

Choice of ξ: for $\sqrt{\delta} = \frac{\epsilon_g}{\mathrm{Out}^{1/2} + 2\epsilon_f^{1/2}} \approx \frac{\epsilon_g}{\mathrm{Out}^{1/2}(\mathcal{P})}$, we get

$$\xi \approx \frac{2\,\epsilon_f\,\epsilon_g^2}{L(\phi)\mathrm{Out}(\mathcal{P})}.$$

Hence, the total number of iterations for getting (ϵ_f, ϵ_g)-solution

Complexity of FGM

Choice of ξ: for $\sqrt{\delta} = \frac{\epsilon_g}{\mathrm{Out}^{1/2} + 2\epsilon_f^{1/2}} \approx \frac{\epsilon_g}{\mathrm{Out}^{1/2}(\mathcal{P})}$, we get

$$\xi \approx \frac{2\,\epsilon_f\,\epsilon_g^2}{L(\phi)\mathrm{Out}(\mathcal{P})}.$$

Hence, the total number of iterations for getting (ϵ_f, ϵ_g)-solution is of the order

$$O\left(\frac{L^{1/2}(\phi)}{\epsilon_g}\mathrm{Out}^{1/2}(\mathcal{P}) \ln \frac{L(\phi)\mathrm{Out}(\mathcal{P})}{\epsilon_f\,\epsilon_g^2}\right).$$

Complexity of FGM

Choice of ξ: for $\sqrt{\delta} = \frac{\epsilon_g}{\mathrm{Out}^{1/2} + 2\epsilon_f^{1/2}} \approx \frac{\epsilon_g}{\mathrm{Out}^{1/2}(\mathcal{P})}$, we get

$$\xi \approx \frac{2\,\epsilon_f\,\epsilon_g^2}{L(\phi)\mathrm{Out}(\mathcal{P})}.$$

Hence, the total number of iterations for getting (ϵ_f, ϵ_g)-solution is of the order

$$O\left(\frac{L^{1/2}(\phi)}{\epsilon_g}\mathrm{Out}^{1/2}(\mathcal{P}) \ln \frac{L(\phi)\mathrm{Out}(\mathcal{P})}{\epsilon_f\,\epsilon_g^2}\right).$$

This is approximately $\sqrt{\cdot}$ of that for GM.

Complexity of FGM

Choice of ξ: for $\sqrt{\delta} = \frac{\epsilon_g}{\mathrm{Out}^{1/2} + 2\epsilon_f^{1/2}} \approx \frac{\epsilon_g}{\mathrm{Out}^{1/2}(\mathcal{P})}$, we get

$$\xi \approx \frac{2\,\epsilon_f\,\epsilon_g^2}{L(\phi)\mathrm{Out}(\mathcal{P})}.$$

Hence, the total number of iterations for getting (ϵ_f, ϵ_g)-solution is of the order

$$O\left(\frac{L^{1/2}(\phi)}{\epsilon_g}\mathrm{Out}^{1/2}(\mathcal{P})\ln\frac{L(\phi)\mathrm{Out}(\mathcal{P})}{\epsilon_f\,\epsilon_g^2}\right).$$

This is approximately $\sqrt{\cdot}$ of that for GM.

NB: dependence on ϵ_f is very weak.

Complexity of FGM

Choice of ξ: for $\sqrt{\delta} = \frac{\epsilon_g}{\mathrm{Out}^{1/2} + 2\epsilon_f^{1/2}} \approx \frac{\epsilon_g}{\mathrm{Out}^{1/2}(\mathcal{P})}$, we get

$$\xi \approx \frac{2\,\epsilon_f\,\epsilon_g^2}{L(\phi)\mathrm{Out}(\mathcal{P})}.$$

Hence, the total number of iterations for getting (ϵ_f, ϵ_g)-solution is of the order

$$O\left(\frac{L^{1/2}(\phi)}{\epsilon_g}\mathrm{Out}^{1/2}(\mathcal{P})\ln\frac{L(\phi)\mathrm{Out}(\mathcal{P})}{\epsilon_f\,\epsilon_g^2}\right).$$

This is approximately $\sqrt{\cdot}$ of that for GM.

NB: dependence on ϵ_f is very weak.

Complexity of objective function enters by $L(\phi) = \frac{1}{\sigma(f)}\|A\|^2$,

Complexity of FGM

Choice of ξ: for $\sqrt{\delta} = \frac{\epsilon_g}{\mathrm{Out}^{1/2} + 2\epsilon_f^{1/2}} \approx \frac{\epsilon_g}{\mathrm{Out}^{1/2}(\mathcal{P})}$, we get

$$\xi \approx \frac{2\,\epsilon_f\,\epsilon_g^2}{L(\phi)\mathrm{Out}(\mathcal{P})}.$$

Hence, the total number of iterations for getting (ϵ_f, ϵ_g)-solution is of the order

$$O\left(\frac{L^{1/2}(\phi)}{\epsilon_g}\mathrm{Out}^{1/2}(\mathcal{P}) \ln \frac{L(\phi)\mathrm{Out}(\mathcal{P})}{\epsilon_f\,\epsilon_g^2}\right).$$

This is approximately $\sqrt{\cdot}$ of that for GM.

NB: dependence on ϵ_f is very weak.

Complexity of objective function enters by $L(\phi) = \frac{1}{\sigma(f)}\|A\|^2$, and $\mathrm{Out}(\mathcal{P})$.

Example 1. Solving system of linear inequalities

Example 1. Solving system of linear inequalities

Problem: Find x : $\|x\|_\infty \leq 1$ and $Ax = b$.

Example 1. Solving system of linear inequalities

Problem: Find $x : \ \|x\|_\infty \leq 1$ and $Ax = b$.

Assumption:
For some $\hat{\epsilon} > 0$ there exist $\hat{x} : \ \|\hat{x}\|_\infty \leq 1 - \hat{\epsilon}$ and $A\hat{x} = b$.

Example 1. Solving system of linear inequalities

Problem: Find $x : \; \|x\|_\infty \le 1$ and $Ax = b$.

Assumption:
For some $\hat{\epsilon} > 0$ there exist $\hat{x} : \; \|\hat{x}\|_\infty \le 1 - \hat{\epsilon}$ and $A\hat{x} = b$.

New problem: $$f^* = \min_{Ax=b} \left[F(x) \stackrel{\rm def}{=} -\sum_{i=1}^{n} \ln \left(1 - (x^{(i)})^2 \right) \right].$$

Example 1. Solving system of linear inequalities

Problem: Find $x: \ \|x\|_\infty \leq 1$ and $Ax = b$.

Assumption:
For some $\hat{\epsilon} > 0$ there exist $\hat{x}: \ \|\hat{x}\|_\infty \leq 1 - \hat{\epsilon}$ and $A\hat{x} = b$.

New problem: $$f^* = \min_{Ax=b} \left[F(x) \stackrel{\text{def}}{=} -\sum_{i=1}^{n} \ln\left(1 - (x^{(i)})^2\right) \right].$$

Dual problem:

Example 1. Solving system of linear inequalities

Problem: Find $x: \ \|x\|_\infty \leq 1$ and $Ax = b$.

Assumption:
For some $\hat{\epsilon} > 0$ there exist $\hat{x}: \ \|\hat{x}\|_\infty \leq 1 - \hat{\epsilon}$ and $A\hat{x} = b$.

New problem: $f^* = \min\limits_{Ax=b} \left[F(x) \stackrel{\mathrm{def}}{=} -\sum\limits_{i=1}^{n} \ln\left(1 - (x^{(i)})^2\right) \right]$.

Dual problem: $\phi(y) = \min\limits_{x} [\langle y, Ax - b\rangle + F(x)]$

Example 1. Solving system of linear inequalities

Problem: Find $x : \ \|x\|_\infty \leq 1$ and $Ax = b$.

Assumption:
For some $\hat{\epsilon} > 0$ there exist $\hat{x} : \ \|\hat{x}\|_\infty \leq 1 - \hat{\epsilon}$ and $A\hat{x} = b$.

New problem: $f^* = \min\limits_{Ax=b} \left[F(x) \stackrel{\mathrm{def}}{=} -\sum\limits_{i=1}^n \ln \left(1 - (x^{(i)})^2 \right) \right]$.

Dual problem: $\phi(y) = \min\limits_x [\langle y, Ax - b\rangle + F(x)]$

$$= -\langle b, y\rangle - \sum_{i=1}^n \left[\sqrt{1 + \langle a_i, y\rangle^2} - 1 - \ln \frac{1+\sqrt{1+\langle a_i,y\rangle^2}}{2} \right] \ \to \ \max_y.$$

Example 1. Solving system of linear inequalities

Problem: Find $x: \ \|x\|_\infty \leq 1$ and $Ax = b$.

Assumption:
For some $\hat{\epsilon} > 0$ there exist $\hat{x}: \ \|\hat{x}\|_\infty \leq 1 - \hat{\epsilon}$ and $A\hat{x} = b$.

New problem: $$f^* = \min_{Ax=b} \left[F(x) \stackrel{\text{def}}{=} -\sum_{i=1}^n \ln \left(1 - (x^{(i)})^2\right) \right].$$

Dual problem: $$\phi(y) = \min_x [\langle y, Ax - b\rangle + F(x)]$$

$$= -\langle b, y\rangle - \sum_{i=1}^n \left[\sqrt{1 + \langle a_i, y\rangle^2} - 1 - \ln \frac{1+\sqrt{1+\langle a_i,y\rangle^2}}{2} \right] \to \max_y .$$

Complexity:

Example 1. Solving system of linear inequalities

Problem: Find $x: \ \|x\|_\infty \le 1$ and $Ax = b$.

Assumption:
For some $\hat\epsilon > 0$ there exist $\hat x: \ \|\hat x\|_\infty \le 1 - \hat\epsilon$ and $A\hat x = b$.

New problem: $$f^* = \min_{Ax=b} \left[F(x) \stackrel{\rm def}{=} -\sum_{i=1}^n \ln\left(1 - (x^{(i)})^2\right)\right].$$

Dual problem: $$\phi(y) = \min_x [\langle y, Ax - b\rangle + F(x)]$$

$$= -\langle b, y\rangle - \sum_{i=1}^n \left[\sqrt{1 + \langle a_i, y\rangle^2} - 1 - \ln \frac{1+\sqrt{1+\langle a_i,y\rangle^2}}{2}\right] \to \max_y.$$

Complexity:

- **IPM**(Newton): $O(\mathrm{Out}(\mathcal{P}))$ Newton iterations.

Example 1. Solving system of linear inequalities

Problem: Find $x: \; \|x\|_\infty \leq 1$ and $Ax = b$.

Assumption:
For some $\hat{\epsilon} > 0$ there exist $\hat{x}: \; \|\hat{x}\|_\infty \leq 1 - \hat{\epsilon}$ and $A\hat{x} = b$.

New problem: $f^* = \min\limits_{Ax=b} \left[F(x) \stackrel{\text{def}}{=} - \sum\limits_{i=1}^{n} \ln \left(1 - (x^{(i)})^2 \right) \right]$.

Dual problem: $\phi(y) = \min\limits_{x} [\langle y, Ax - b \rangle + F(x)]$

$$= -\langle b, y \rangle - \sum_{i=1}^{n} \left[\sqrt{1 + \langle a_i, y \rangle^2} - 1 - \ln \frac{1 + \sqrt{1 + \langle a_i, y \rangle^2}}{2} \right] \rightarrow \max_y .$$

Complexity:

- **IPM**(Newton): $O(\mathrm{Out}(\mathcal{P}))$ Newton iterations.
- **FGM**: $\frac{1}{\epsilon_g} \|A\| \mathrm{Out}^{1/2}(\mathcal{P})$ gradient iterations.

Example 1. Solving system of linear inequalities

Problem: Find $x: \ \|x\|_\infty \leq 1$ and $Ax = b$.

Assumption:
For some $\hat{\epsilon} > 0$ there exist $\hat{x}: \ \|\hat{x}\|_\infty \leq 1 - \hat{\epsilon}$ and $A\hat{x} = b$.

New problem: $f^* = \min\limits_{Ax=b} \left[F(x) \stackrel{\mathrm{def}}{=} -\sum\limits_{i=1}^n \ln\left(1 - (x^{(i)})^2\right) \right].$

Dual problem: $\phi(y) = \min\limits_x [\langle y, Ax - b\rangle + F(x)]$

$$= -\langle b, y\rangle - \sum_{i=1}^n \left[\sqrt{1 + \langle a_i, y\rangle^2} - 1 - \ln \frac{1+\sqrt{1+\langle a_i,y\rangle^2}}{2} \right] \to \max_y.$$

Complexity:

- **IPM**(Newton): $O(\mathrm{Out}(\mathcal{P}))$ Newton iterations.
- **FGM**: $\frac{1}{\epsilon_g}\|A\|\mathrm{Out}^{1/2}(\mathcal{P})$ gradient iterations.

NB: $\mathrm{Out}(\mathcal{P}) \leq n \ln \frac{1}{\hat{\epsilon}}$.

Example 2. Entropy projection

Example 2. Entropy projection

Problem: $$\min_{x \in \Delta_n} \left\{ \eta(x) \stackrel{\text{def}}{=} \sum_{i=1}^{n} x^{(i)} \ln x^{(i)} : \; Ax = 0 \right\}.$$

Example 2. Entropy projection

Problem: $$\min_{x \in \Delta_n} \left\{ \eta(x) \stackrel{\text{def}}{=} \sum_{i=1}^{n} x^{(i)} \ln x^{(i)} : \; Ax = 0 \right\}.$$

NB: $\text{Out}(\mathcal{P}) = \ln n$.

Example 2. Entropy projection

Problem: $$\min_{x \in \Delta_n} \left\{ \eta(x) \stackrel{\text{def}}{=} \sum_{i=1}^{n} x^{(i)} \ln x^{(i)} : \; Ax = 0 \right\}.$$

NB: $\text{Out}(\mathcal{P}) = \ln n$.

Dual problem:

$$\phi(y) = \min_{x \in \Delta_n} [-\langle y, Ax \rangle + \eta(x)]$$

Example 2. Entropy projection

Problem: $\min\limits_{x \in \Delta_n} \left\{ \eta(x) \stackrel{\mathrm{def}}{=} \sum\limits_{i=1}^n x^{(i)} \ln x^{(i)} : \; Ax = 0 \right\}.$

NB: $\mathrm{Out}(\mathcal{P}) = \ln n.$

Dual problem:

$$\phi(y) = \min_{x \in \Delta_n} [-\langle y, Ax \rangle + \eta(x)] = -\ln \sum_{i=1}^n e^{\langle a_i, y \rangle}$$

Example 2. Entropy projection

Problem: $$\min_{x \in \Delta_n} \left\{ \eta(x) \stackrel{\text{def}}{=} \sum_{i=1}^{n} x^{(i)} \ln x^{(i)} : \ Ax = 0 \right\}.$$

NB: $\mathrm{Out}(\mathcal{P}) = \ln n$.

Dual problem:

$$\phi(y) = \min_{x \in \Delta_n} [-\langle y, Ax \rangle + \eta(x)] = -\ln \sum_{i=1}^{n} e^{\langle a_i, y \rangle} \ \rightarrow \ \max_{y \in R^m}.$$

Example 2. Entropy projection

Problem: $\min\limits_{x \in \Delta_n} \left\{ \eta(x) \stackrel{\mathrm{def}}{=} \sum\limits_{i=1}^{n} x^{(i)} \ln x^{(i)} : \; Ax = 0 \right\}.$

NB: $\mathrm{Out}(\mathcal{P}) = \ln n.$

Dual problem:

$$\phi(y) = \min_{x \in \Delta_n} [-\langle y, Ax \rangle + \eta(x)] = -\ln \sum_{i=1}^{n} e^{\langle a_i, y \rangle} \; \to \; \max_{y \in R^m}.$$

Optimal solution y^* can be very big (or do not exist).

Example 2. Entropy projection

Problem: $\min\limits_{x \in \Delta_n} \left\{ \eta(x) \stackrel{\text{def}}{=} \sum\limits_{i=1}^{n} x^{(i)} \ln x^{(i)} : \; Ax = 0 \right\}.$

NB: $\mathrm{Out}(\mathcal{P}) = \ln n.$

Dual problem:

$$\phi(y) = \min_{x \in \Delta_n} [-\langle y, Ax \rangle + \eta(x)] = -\ln \sum_{i=1}^{n} e^{\langle a_i, y \rangle} \; \to \; \max_{y \in R^m}.$$

Optimal solution y^* can be very big (or do not exist).

Complexity of dual FGM:

Example 2. Entropy projection

Problem: $\min\limits_{x \in \Delta_n} \left\{ \eta(x) \stackrel{\mathrm{def}}{=} \sum\limits_{i=1}^{n} x^{(i)} \ln x^{(i)} : \; Ax = 0 \right\}$.

NB: $\mathrm{Out}(\mathcal{P}) = \ln n$.

Dual problem:

$$\phi(y) = \min_{x \in \Delta_n} [-\langle y, Ax \rangle + \eta(x)] = -\ln \sum_{i=1}^{n} e^{\langle a_i, y \rangle} \;\to\; \max_{y \in R^m}.$$

Optimal solution y^* can be very big (or do not exist).

Complexity of dual FGM:

$$O\left(\frac{\ln^{1/2} n}{\epsilon_g} \max_{1 \leq i \leq n} \|a_i\|_2 \right) \text{ gradient iterations.}$$

Conclusion

Conclusion

- Our complexity bounds depend on ϵ_g in an optimal way.

Conclusion

- Our complexity bounds depend on ϵ_g in an optimal way.
- They almost do not depend on ϵ_f.

Conclusion

- Our complexity bounds depend on ϵ_g in an optimal way.
- They almost do not depend on ϵ_f.
- This is natural since we "eliminate" f by dualization.

Conclusion

- Our complexity bounds depend on ϵ_g in an optimal way.
- They almost do not depend on ϵ_f.
- This is natural since we "eliminate" f by dualization.

THANK YOU FOR YOUR ATTENTION!