Golden Ratio Algorithms for Variational Inequalities

Yura Malitsky

Georg-August-University Göttingen
Institute for Numerical and Applied Mathematics

Games, Dynamics and Optimization
Cluj-Napoca, 2019

First-order method for general convex problems

- Lipschitz constants are bad;
- Lipschitz assumptions are worse;
- Linesearch are ugly.

Introduction

Variational inequality problem (VI):

find $x^* \in X = \mathbb{R}^d$ such that

$$\langle F(x^*), x - x^* \rangle + g(x) - g(x^*) \geq 0 \quad \forall x \in X,$$

Variational inequality problem (VI):

find $x^* \in X = \mathbb{R}^d$ such that

$$\langle F(x^*), x - x^* \rangle + g(x) - g(x^*) \geq 0 \quad \forall x \in X,$$

where

- $F\colon X \to X$ is monotone: $\langle F(u) - F(v), u - v \rangle \geq 0 \quad \forall u, v$
- $g\colon X \to (-\infty, +\infty]$ is a proper lsc convex function

Variational inequality problem (VI):

find $x^* \in X = \mathbb{R}^d$ such that

$$\langle F(x^*), x - x^* \rangle + g(x) - g(x^*) \geq 0 \quad \forall x \in X,$$

where

- $F\colon X \to X$ is monotone: $\langle F(u) - F(v), u - v \rangle \geq 0 \quad \forall u, v$
- $g\colon X \to (-\infty, +\infty]$ is a proper lsc convex function

VI as a monotone operator inclusion:

$$0 \in F(x^*) + \partial g(x^*)$$

Motivation–1

Composite minimization:

$$\min_x f(x) + g(x)$$

- $f \colon X \to \mathbb{R}$ is a convex smooth function with Lipschitz ∇f
- $g \colon X \to (-\infty, +\infty]$ is a proper lsc convex function

Motivation–1

Composite minimization:

$$\min_x f(x) + g(x)$$

- $f\colon X \to \mathbb{R}$ is a convex smooth function with Lipschitz ∇f
- $g\colon X \to (-\infty, +\infty]$ is a proper lsc convex function

First-order optimality condition:

$$\langle \nabla f(x^*), x - x^* \rangle + g(x) - g(x^*) \geq 0 \quad \forall x \in X.$$

Motivation–2

Saddle point problem:

$$\min_x \max_y \mathcal{L}(x, y) := g_1(x) + K(x, y) - g_2^*(y)$$

- $K\colon X \times Y \to \mathbb{R}$ is smooth convex-concave
- $g_1\colon X \to (-\infty, +\infty]$, $g_2\colon Y \to (-\infty, +\infty]$ are convex lsc

Motivation–2

Saddle point problem:

$$\min_x \max_y \mathcal{L}(x, y) := g_1(x) + K(x, y) - g_2^*(y)$$

- $K \colon X \times Y \to \mathbb{R}$ is smooth convex-concave
- $g_1 \colon X \to (-\infty, +\infty]$, $g_2 \colon Y \to (-\infty, +\infty]$ are convex lsc

Let

$$z = \begin{pmatrix} x \\ y \end{pmatrix}, \quad F(z) = \begin{pmatrix} \nabla_x K(x, y) \\ -\nabla_y K(x, y) \end{pmatrix}, \quad g(z) = g_1(x) + g_2^*(y)$$

Motivation–2

Saddle point problem:

$$\min_x \max_y \mathcal{L}(x, y) := g_1(x) + K(x, y) - g_2^*(y)$$

- $K\colon X \times Y \to \mathbb{R}$ is smooth convex-concave
- $g_1\colon X \to (-\infty, +\infty]$, $g_2\colon Y \to (-\infty, +\infty]$ are convex lsc

Let

$$z = \begin{pmatrix} x \\ y \end{pmatrix}, \quad F(z) = \begin{pmatrix} \nabla_x K(x, y) \\ -\nabla_y K(x, y) \end{pmatrix}, \quad g(z) = g_1(x) + g_2^*(y)$$

First-order optimality condition:

$$\langle F(z^*), z - z^* \rangle + g(z) - g(z^*) \geq 0 \quad \forall z \in X \times Y$$

Brief overview

$$\langle F(x^*), x - x^* \rangle + g(x) - g(x^*) \geq 0 \quad \forall x \in X,$$

Brief overview

$$\langle F(x^*), x - x^* \rangle + g(x) - g(x^*) \geq 0 \quad \forall x \in X,$$

Backward method (PPA):

$$x^{k+1} = J_{F+\partial g} x^k$$

Brief overview

$$\langle F(x^*), x - x^* \rangle + g(x) - g(x^*) \geq 0 \quad \forall x \in X,$$

Backward method (PPA):

$$x^{k+1} = J_{F+\partial g} x^k$$

Forward-backward method:

$$x^{k+1} = \mathrm{prox}_{\lambda g}(x^k - \lambda F(x^k))$$

Brief overview

$$\langle F(x^*), x - x^* \rangle + g(x) - g(x^*) \geq 0 \quad \forall x \in X,$$

Backward method (PPA):

$$x^{k+1} = J_{F+\partial g} x^k$$

Forward-backward method:

$$x^{k+1} = \mathrm{prox}_{\lambda g}(x^k - \lambda F(x^k))$$

doesn't work...

$$\langle F(x^*), x - x^* \rangle + g(x) - g(x^*) \geq 0 \quad \forall x \in X,$$

Backward method (PPA):

$$x^{k+1} = J_{F+\partial g} x^k$$

Forward-backward method:

$$x^{k+1} = \text{prox}_{\lambda g}(x^k - \lambda F(x^k))$$

doesn't work... Try: change x^k or $F(x^k)$.

$$\langle F(x^*), x - x^* \rangle + g(x) - g(x^*) \geq 0 \quad \forall x \in X,$$

Backward method (PPA):

$$x^{k+1} = J_{F+\partial g} x^k$$

Forward-backward method:

$$x^{k+1} = \operatorname{prox}_{\lambda g}(x^k - \lambda F(x^k))$$

doesn't work... Try: change x^k or $F(x^k)$.

Extragradient method [Korpelevich '76]:

Brief overview

$$\langle F(x^*), x - x^* \rangle + g(x) - g(x^*) \geq 0 \quad \forall x \in X,$$

Backward method (PPA):

$$x^{k+1} = J_{F+\partial g} x^k$$

Forward-backward method:

$$x^{k+1} = \mathrm{prox}_{\lambda g}(x^k - \lambda F(x^k))$$

doesn't work... Try: change x^k or $F(x^k)$.

Extragradient method [Korpelevich '76]:

$$x^{k+1} = \mathrm{prox}_{\lambda g}(x^k - \lambda F(y^k))$$

Brief overview

$$\langle F(x^*), x - x^* \rangle + g(x) - g(x^*) \geq 0 \quad \forall x \in X,$$

Backward method (PPA):

$$x^{k+1} = J_{F+\partial g} x^k$$

Forward-backward method:

$$x^{k+1} = \text{prox}_{\lambda g}(x^k - \lambda F(x^k))$$

doesn't work... Try: change x^k or $F(x^k)$.

Extragradient method [Korpelevich '76]:

$$y^k = \text{prox}_{\lambda g}(x^k - \lambda F(x^k))$$
$$x^{k+1} = \text{prox}_{\lambda g}(x^k - \lambda F(y^k))$$

Simpler modifications of FB

Forward-backward method:

$$x^{k+1} = \text{prox}_{\lambda g}(x^k - \lambda F(x^k))$$

doesn't work...

Simpler modifications of FB

Forward-backward method:

$$x^{k+1} = \operatorname{prox}_{\lambda g}(x^k - \lambda F(x^k))$$

doesn't work...

- Change $F(x^k)$:

Simpler modifications of FB

Forward-backward method:

$$x^{k+1} = \operatorname{prox}_{\lambda g}(x^k - \lambda F(x^k))$$

doesn't work...

- Change $F(x^k)$:

$$x^{k+1} = \operatorname{prox}_{\lambda g}(x^k - \lambda F(2x^k - x^{k-1}))$$

Simpler modifications of FB

Forward-backward method:

$$x^{k+1} = \operatorname{prox}_{\lambda g}(x^k - \lambda F(x^k))$$

doesn't work...

- Change $F(x^k)$:

$$x^{k+1} = \operatorname{prox}_{\lambda g}(x^k - \lambda F(2x^k - x^{k-1}))$$

- Change $F(x^k)$:

Simpler modifications of FB

Forward-backward method:

$$x^{k+1} = \text{prox}_{\lambda g}(x^k - \lambda F(x^k))$$

doesn't work...

- Change $F(x^k)$:

$$x^{k+1} = \text{prox}_{\lambda g}(x^k - \lambda F(2x^k - x^{k-1}))$$

- Change $F(x^k)$:

$$x^{k+1} = \text{prox}_{\lambda g}\Big(x^k - \lambda(2F(x^k) - F(x^{k-1}))\Big)$$

Simpler modifications of FB

Forward-backward method:

$$x^{k+1} = \text{prox}_{\lambda g}(x^k - \lambda F(x^k))$$

doesn't work...

- Change $F(x^k)$:

$$x^{k+1} = \text{prox}_{\lambda g}(x^k - \lambda F(2x^k - x^{k-1}))$$

- Change $F(x^k)$:

$$x^{k+1} = \text{prox}_{\lambda g}\Big(x^k - \lambda(2F(x^k) - F(x^{k-1}))\Big)$$

- Change x^k?

Simpler modifications of FB

Forward-backward method:

$$x^{k+1} = \mathrm{prox}_{\lambda g}(x^k - \lambda F(x^k))$$

doesn't work...

- Change $F(x^k)$:

$$x^{k+1} = \mathrm{prox}_{\lambda g}(x^k - \lambda F(2x^k - x^{k-1}))$$

- Change $F(x^k)$:

$$x^{k+1} = \mathrm{prox}_{\lambda g}\Big(x^k - \lambda(2F(x^k) - F(x^{k-1}))\Big)$$

- Change x^k?

$$x^{k+1} = \mathrm{prox}_{\lambda g}(? - \lambda F(x^k))$$

Golden Ratio Algorithm

$$x^{k+1} = \operatorname{prox}_{\lambda g}\big(\bar{x}^k - \lambda F(x^k)\big)$$

Golden Ratio Algorithm

$$x^{k+1} = \text{prox}_{\lambda g}(\bar{x}^k - \lambda F(x^k))$$

Can we take $\bar{x}^k = x^k + \alpha_k(x^k - x^{k-1})$?

Golden Ratio Algorithm

$$x^{k+1} = \text{prox}_{\lambda g}(\bar{x}^k - \lambda F(x^k))$$

Can we take $\bar{x}^k = x^k + \alpha_k(x^k - x^{k-1})$?   No way!

Golden Ratio Algorithm

$$x^{k+1} = \mathrm{prox}_{\lambda g}(\bar{x}^k - \lambda F(x^k))$$

Can we take $\bar{x}^k = x^k + \alpha_k(x^k - x^{k-1})$? No way!

Can we take $\bar{x}^k = x^k - \alpha_k(x^k - x^{k-1})$? Almost...

Golden Ratio Algorithm

$$x^{k+1} = \text{prox}_{\lambda g}(\bar{x}^k - \lambda F(x^k))$$

Can we take $\bar{x}^k = x^k + \alpha_k(x^k - x^{k-1})$? No way!

Can we take $\bar{x}^k = x^k - \alpha_k(x^k - x^{k-1})$? Almost...

Idea: $\bar{x}^k \in \text{conv}\{x^k, x^{k-1}, \dots, x^0\}$. Let $\varphi = \frac{\sqrt{5}+1}{2} = 1.618\dots$

Golden Ratio Algorithm

$$x^{k+1} = \mathrm{prox}_{\lambda g}(\bar{x}^k - \lambda F(x^k))$$

Can we take $\bar{x}^k = x^k + \alpha_k(x^k - x^{k-1})$? No way!

Can we take $\bar{x}^k = x^k - \alpha_k(x^k - x^{k-1})$? Almost...

Idea: $\bar{x}^k \in \mathrm{conv}\{x^k, x^{k-1}, \dots, x^0\}$. Let $\varphi = \frac{\sqrt{5}+1}{2} = 1.618\dots$

Golden Ratio Algorithm (GRAAL):

$$\bar{x}^k = \frac{(\varphi - 1)x^k + \bar{x}^{k-1}}{\varphi}$$

Golden Ratio Algorithm

$$x^{k+1} = \operatorname{prox}_{\lambda g}(\bar{x}^k - \lambda F(x^k))$$

Can we take $\bar{x}^k = x^k + \alpha_k(x^k - x^{k-1})$? No way!

Can we take $\bar{x}^k = x^k - \alpha_k(x^k - x^{k-1})$? Almost...

Idea: $\bar{x}^k \in \operatorname{conv}\{x^k, x^{k-1}, \dots, x^0\}$. Let $\varphi = \frac{\sqrt{5}+1}{2} = 1.618\dots$

Golden Ratio Algorithm (GRAAL):

$$\bar{x}^k = \frac{(\varphi - 1)x^k + \bar{x}^{k-1}}{\varphi}$$
$$x^{k+1} = \operatorname{prox}_{\lambda g}(\bar{x}^k - \lambda F(x^k))$$

Golden Ratio Algorithm

$$x^{k+1} = \operatorname{prox}_{\lambda g}(\bar{x}^k - \lambda F(x^k))$$

Can we take $\bar{x}^k = x^k + \alpha_k(x^k - x^{k-1})$? No way!

Can we take $\bar{x}^k = x^k - \alpha_k(x^k - x^{k-1})$? Almost...

Idea: $\bar{x}^k \in \operatorname{conv}\{x^k, x^{k-1}, \dots, x^0\}$. Let $\varphi = \frac{\sqrt{5}+1}{2} = 1.618\dots$

Golden Ratio Algorithm (GRAAL):

$$\bar{x}^k = \frac{(\varphi - 1)x^k + \bar{x}^{k-1}}{\varphi}$$

$$x^{k+1} = \operatorname{prox}_{\lambda g}(\bar{x}^k - \lambda F(x^k))$$

Theorem

Let $F\colon X \to X$ is monotone and L–Lipschitz, g is convex lsc. Then (x^k) converges to a solution of VI.

Difficult VI

Composite minimization:

$$\min_x f(x) + g(x),$$

where f, g are convex lsc, g is prox-friendly, f is smooth, but

Difficult VI

Composite minimization:

$$\min_x f(x) + g(x),$$

where f, g are convex lsc, g is prox-friendly, f is smooth, but

- either ∇f is L–Lipschitz but L is unknown;
- or ∇f is not Lipschitz.

Difficult VI

Composite minimization:

$$\min_x f(x) + g(x),$$

where f, g are convex lsc, g is prox-friendly, f is smooth, but

- either ∇f is L–Lipschitz but L is unknown;
- or ∇f is not Lipschitz.

Constrained minimization:

$$\min f(x) \quad \text{s.t.} \quad h(x) \leq 0,$$

where f, h are convex smooth with Lipschitz ∇f, ∇h.

Difficult VI

Composite minimization:

$$\min_x f(x) + g(x),$$

where f, g are convex lsc, g is prox-friendly, f is smooth, but

- either ∇f is L–Lipschitz but L is unknown;
- or ∇f is not Lipschitz.

Constrained minimization:

$$\min f(x) \quad \text{s.t.} \quad h(x) \le 0,$$

where f, h are convex smooth with Lipschitz ∇f, ∇h.

Minmax form: $\min_x \max_y f(x) + yh(x) - \delta_{\ge 0}(y)$

Difficult VI

Composite minimization:

$$\min_x f(x) + g(x),$$

where f, g are convex lsc, g is prox-friendly, f is smooth, but

- either ∇f is L–Lipschitz but L is unknown;
- or ∇f is not Lipschitz.

Constrained minimization:

$$\min f(x) \quad \text{s.t.} \quad h(x) \leq 0,$$

where f, h are convex smooth with Lipschitz ∇f, ∇h.

Minmax form: $\min_x \max_y f(x) + yh(x) - \delta_{\geq 0}(y)$

$$F\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \nabla f(x) + y\nabla h(x) \\ -h(x) \end{pmatrix}$$

Difficult VI

Composite minimization:

$$\min_x f(x) + g(x),$$

where f, g are convex lsc, g is prox-friendly, f is smooth, but

- either ∇f is L–Lipschitz but L is unknown;
- or ∇f is not Lipschitz.

Constrained minimization:

$$\min f(x) \quad \text{s.t.} \quad h(x) \leq 0,$$

where f, h are convex smooth with Lipschitz ∇f, ∇h.

Minmax form: $\min_x \max_y f(x) + yh(x) - \delta_{\geq 0}(y)$

$$F\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \nabla f(x) + y \nabla h(x) \\ -h(x) \end{pmatrix} \text{ — unlikely to be Lipschitz}$$

Adaptive method

Golden ratio algorithm:

$$\bar{x}^k = \frac{(\varphi - 1)x^k + \bar{x}^{k-1}}{\varphi}$$
$$x^{k+1} = \operatorname{prox}_{\lambda g}(\bar{x}^k - \lambda F(x^k))$$

Adaptive method

Golden ratio algorithm:

$$\bar{x}^k = \frac{(\varphi - 1)x^k + \bar{x}^{k-1}}{\varphi}$$
$$x^{k+1} = \operatorname{prox}_{\lambda g}(\bar{x}^k - \lambda F(x^k))$$

$$\lambda_k \approx \frac{\|x^k - x^{k-1}\|}{\|F(x^k) - F(x^{k-1})\|}$$

Adaptive method

Golden ratio algorithm:

$$\bar{x}^k = \frac{(\varphi - 1)x^k + \bar{x}^{k-1}}{\varphi}$$
$$x^{k+1} = \text{prox}_{\lambda g}(\bar{x}^k - \lambda F(x^k))$$

$$\lambda_k \approx \frac{\|x^k - x^{k-1}\|}{\|F(x^k) - F(x^{k-1})\|}$$

Adaptive golden ratio algorithm:

$$\lambda_k = \qquad \frac{9}{16\lambda_{k-2}} \frac{\|x^k - x^{k-1}\|^2}{\|F(x^k) - F(x^{k-1})\|^2}$$

Adaptive method

Golden ratio algorithm:

$$\bar{x}^k = \frac{(\varphi - 1)x^k + \bar{x}^{k-1}}{\varphi}$$

$$x^{k+1} = \text{prox}_{\lambda g}(\bar{x}^k - \lambda F(x^k))$$

$$\lambda_k \approx \frac{\|x^k - x^{k-1}\|}{\|F(x^k) - F(x^{k-1})\|}$$

Adaptive golden ratio algorithm:

$$\lambda_k = \qquad \frac{10}{9}\lambda_{k-1}, \frac{9}{16\lambda_{k-2}} \frac{\|x^k - x^{k-1}\|^2}{\|F(x^k) - F(x^{k-1})\|^2}$$

Adaptive method

Golden ratio algorithm:

$$\bar{x}^k = \frac{(\varphi - 1)x^k + \bar{x}^{k-1}}{\varphi}$$
$$x^{k+1} = \text{prox}_{\lambda g}(\bar{x}^k - \lambda F(x^k))$$

$$\lambda_k \approx \frac{\|x^k - x^{k-1}\|}{\|F(x^k) - F(x^{k-1})\|}$$

Adaptive golden ratio algorithm:

$$\lambda_k = \min\Big\{\frac{10}{9}\lambda_{k-1}, \frac{9}{16\lambda_{k-2}}\frac{\|x^k - x^{k-1}\|^2}{\|F(x^k) - F(x^{k-1})\|^2}\Big\}$$

Adaptive method

Golden ratio algorithm:

$$\bar{x}^k = \frac{(\varphi - 1)x^k + \bar{x}^{k-1}}{\varphi}$$
$$x^{k+1} = \mathrm{prox}_{\lambda g}(\bar{x}^k - \lambda F(x^k))$$

$$\lambda_k \approx \frac{\|x^k - x^{k-1}\|}{\|F(x^k) - F(x^{k-1})\|}$$

Adaptive golden ratio algorithm:

$$\lambda_k = \min\Big\{\frac{10}{9}\lambda_{k-1}, \frac{9}{16\lambda_{k-2}}\frac{\|x^k - x^{k-1}\|^2}{\|F(x^k) - F(x^{k-1})\|^2}\Big\}$$
$$\bar{x}^k = \frac{x^k + 2\bar{x}^{k-1}}{3}$$
$$x^{k+1} = \mathrm{prox}_{\lambda_k g}(\bar{x}^k - \lambda_k F(x^k))$$

Adaptive method

Adaptive golden ratio algorithm:

$$\lambda_k = \min\Big\{\frac{10}{9}\lambda_{k-1}, \frac{9}{16\lambda_{k-2}}\frac{\|x^k - x^{k-1}\|^2}{\|F(x^k) - F(x^{k-1})\|^2}\Big\}$$

$$\bar{x}^k = \frac{x^k + 2\bar{x}^{k-1}}{3}$$

$$x^{k+1} = \operatorname{prox}_{\lambda_k g}(\bar{x}^k - \lambda_k F(x^k))$$

Theorem

Let $F\colon X \to X$ is monotone and locally Lipschitz, g is convex lsc. Then (x^k) converges to a solution of VI.

Rates

F is loc. Lipschitz, (x^k) is bounded $\Longrightarrow \lambda_k$ is separated from zero.

Rates

F is loc. Lipschitz, (x^k) is bounded $\Longrightarrow \lambda_k$ is separated from zero.

- ergodic $O(1/N)$ rate for the dual gap function

$$e(v) = \max_{u \in \operatorname{dom} g} \Psi(u, v) := \langle F(u), v - u \rangle + g(v) - g(u)$$

Rates

F is loc. Lipschitz, (x^k) is bounded $\implies \lambda_k$ is separated from zero.

- ergodic $O(1/N)$ rate for the dual gap function

$$e(v) = \max_{u \in \operatorname{dom} g} \Psi(u, v) := \langle F(u), v - u \rangle + g(v) - g(u)$$

- linear rate under error bounds

Rates

F is loc. Lipschitz, (x^k) is bounded $\implies \lambda_k$ is separated from zero.

- ergodic $O(1/N)$ rate for the dual gap function

$$e(v) = \max_{u \in \operatorname{dom} g} \Psi(u, v) := \langle F(u), v - u \rangle + g(v) - g(u)$$

- linear rate under error bounds

Rates

F is loc. Lipschitz, (x^k) is bounded $\Longrightarrow \lambda_k$ is separated from zero.

- ergodic $O(1/N)$ rate for the dual gap function

$$e(v) = \max_{u \in \operatorname{dom} g} \Psi(u, v) := \langle F(u), v - u \rangle + g(v) - g(u)$$

- linear rate under error bounds

Composite minimization:

$$\min_x h(x) := f(x) + g(x)$$

Rates

F is loc. Lipschitz, (x^k) is bounded $\Longrightarrow \lambda_k$ is separated from zero.

- ergodic $O(1/N)$ rate for the dual gap function

$$e(v) = \max_{u \in \operatorname{dom} g} \Psi(u, v) := \langle F(u), v - u \rangle + g(v) - g(u)$$

- linear rate under error bounds

Composite minimization:

$$\min_x h(x) := f(x) + g(x)$$

$$\sum_{i=1}^{k} \lambda_i (h(x^i) - h_*) \leq C \Longrightarrow h\Big(\frac{\sum_i \lambda_i x^i}{\sum_i \lambda_i}\Big) - h_* \leq \frac{C}{\sum_i \lambda_i}$$

Sparse logistic regression

$$\min_x h(x) := \sum_{i=1}^m \log(1 + e^{-b_i \langle a_i, x\rangle}) + \gamma \|x\|_1,$$

where $x \in \mathbb{R}^n$, $a_i \in \mathbb{R}^n$, and $b_i \in \{-1, 1\}$, $\gamma > 0$.

Sparse logistic regression

$$\min_x h(x) := \sum_{i=1}^m \log(1 + e^{-b_i \langle a_i, x \rangle}) + \gamma \|x\|_1,$$

where $x \in \mathbb{R}^n$, $a_i \in \mathbb{R}^n$, and $b_i \in \{-1, 1\}$, $\gamma > 0$.

∇f is L–Lipschitz with $L = \frac{1}{4}\|A\|_2^2$

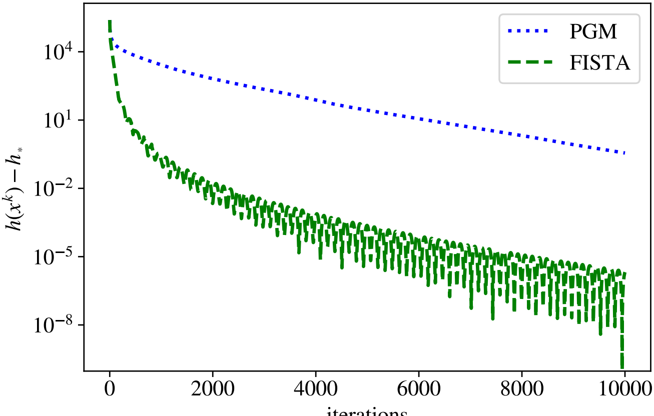
Sparse logistic regression

$$\min_x h(x) := \sum_{i=1}^m \log(1 + e^{-b_i \langle a_i, x \rangle}) + \gamma \|x\|_1,$$

where $x \in \mathbb{R}^n$, $a_i \in \mathbb{R}^n$, and $b_i \in \{-1, 1\}$, $\gamma > 0$.

∇f is L–Lipschitz with $L = \frac{1}{4}\|A\|_2^2$

LIBSVM: *kdda-2010*, $n = 2014669$, $m = 510302$



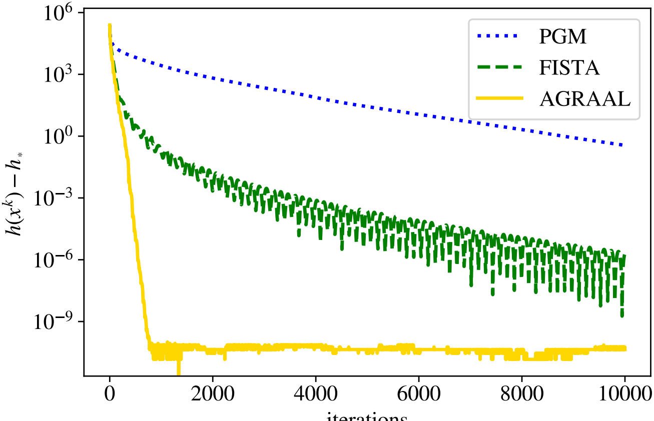
Sparse logistic regression

$$\min_x h(x) := \sum_{i=1}^m \log(1 + e^{-b_i \langle a_i, x \rangle}) + \gamma \|x\|_1,$$

where $x \in \mathbb{R}^n$, $a_i \in \mathbb{R}^n$, and $b_i \in \{-1, 1\}$, $\gamma > 0$.

∇f is L–Lipschitz with $L = \frac{1}{4}\|A\|_2^2$

LIBSVM: *kdda-2010*, $n = 2014669$, $m = 510302$



Beyond monotonicity

$$\langle F(x^*), x - x^* \rangle + g(x) - g(x^*) \geq 0 \quad \forall x \in X,$$

Where monotonicity plays role?

Beyond monotonicity

$$\langle F(x^*), x - x^* \rangle + g(x) - g(x^*) \geq 0 \quad \forall x \in X,$$

Where monotonicity plays role?

$$\begin{aligned}\alpha\|\bar{x}^{k+1} - x\|^2 + \beta_k\|x^{k+1} - x^k\|^2 + \gamma_k\|x^k - \bar{x}^k\|^2 \\ \leq \alpha\|\bar{x}^k - x\|^2 + \beta_{k-1}\|x^k - x^{k-1}\|^2 \\ + 2\lambda_k[\langle F(x^k), x^k - x\rangle + g(x^k) - g(x)]\end{aligned}$$

Beyond monotonicity

$$\langle F(x^*), x - x^* \rangle + g(x) - g(x^*) \geq 0 \quad \forall x \in X,$$

Where monotonicity plays role?

$$\alpha\|\bar{x}^{k+1} - x\|^2 + \beta_k\|x^{k+1} - x^k\|^2 + \gamma_k\|x^k - \bar{x}^k\|^2 \\ \leq \alpha\|\bar{x}^k - x\|^2 + \beta_{k-1}\|x^k - x^{k-1}\|^2 \\ + 2\lambda_k[\langle F(x^k), x^k - x\rangle + g(x^k) - g(x)]$$

If $x = x^* \in S$, then by monotonicity

$$\langle F(x^k), x^k - x^* \rangle + g(x^k) - g(x^*) \geq 0$$

Beyond monotonicity

$$\langle F(x^*), x - x^* \rangle + g(x) - g(x^*) \geq 0 \quad \forall x \in X,$$

Where monotonicity plays role?

$$\begin{aligned}\alpha\|\bar{x}^{k+1} - x\|^2 + \beta_k\|x^{k+1} - x^k\|^2 + \gamma_k\|x^k - \bar{x}^k\|^2 \\ \leq \alpha\|\bar{x}^k - x\|^2 + \beta_{k-1}\|x^k - x^{k-1}\|^2 \\ + 2\lambda_k[\langle F(x^k), x^k - x\rangle + g(x^k) - g(x)]\end{aligned}$$

If $x = x^* \in S$, then by monotonicity

$$\langle F(x^k), x^k - x^* \rangle + g(x^k) - g(x^*) \geq 0$$

$$\exists \tilde{x} \in \operatorname{dom} g \quad \text{s.t.} \quad \langle F(x), x - \tilde{x} \rangle + g(x) - g(\tilde{x}) \geq 0 \quad \forall x$$

- more general than monotonicity or pseudo-monotonicity

Examples

$$\exists \tilde{x} \in \operatorname{dom} g \quad \text{s.t.} \quad \langle F(x), x - \tilde{x} \rangle + g(x) - g(\tilde{x}) \geq 0 \quad \forall x$$

Examples

$$\exists \tilde{x} \in \operatorname{dom} g \quad \text{s.t.} \quad \langle F(x), x - \tilde{x} \rangle + g(x) - g(\tilde{x}) \geq 0 \quad \forall x$$

Example 1

$\min_x f(x) + \|x\|_1$ with $\|\nabla f(x)\|_\infty \leq 1$

Examples

$$\exists \tilde{x} \in \operatorname{dom} g \quad \text{s.t.} \quad \langle F(x), x - \tilde{x} \rangle + g(x) - g(\tilde{x}) \geq 0 \quad \forall x$$

Example 1

$\min_x f(x) + \|x\|_1$ with $\|\nabla f(x)\|_\infty \leq 1$

$$\tilde{x} = 0 \quad \Rightarrow \quad \langle \nabla f(x), x \rangle + \|x\|_1 \geq 0$$

Examples

$$\exists \tilde{x} \in \operatorname{dom} g \quad \text{s.t.} \quad \langle F(x), x - \tilde{x} \rangle + g(x) - g(\tilde{x}) \geq 0 \quad \forall x$$

Example 1

$\min_x f(x) + \|x\|_1$ with $\|\nabla f(x)\|_\infty \leq 1$

$$\tilde{x} = 0 \quad \Rightarrow \quad \langle \nabla f(x), x \rangle + \|x\|_1 \geq 0$$

Example 2

find $x \neq 0$ s.t. $M(x)x = 0$, where M is psd matrix for all x

Examples

$$\exists \tilde{x} \in \operatorname{dom} g \quad \text{s.t.} \quad \langle F(x), x - \tilde{x} \rangle + g(x) - g(\tilde{x}) \geq 0 \quad \forall x$$

Example 1

$\min_x f(x) + \|x\|_1$ with $\|\nabla f(x)\|_\infty \leq 1$

$$\tilde{x} = 0 \quad \Rightarrow \quad \langle \nabla f(x), x \rangle + \|x\|_1 \geq 0$$

Example 2

find $x \neq 0$ s.t. $M(x)x = 0$, where M is psd matrix for all x

$$g \equiv 0, \quad \tilde{x} = 0, \quad F(x) = M(x)x \quad \Rightarrow \quad \langle F(x), x \rangle \geq 0$$

Examples

$$\exists \tilde{x} \in \operatorname{dom} g \quad \text{s.t.} \quad \langle F(x), x - \tilde{x} \rangle + g(x) - g(\tilde{x}) \geq 0 \quad \forall x$$

Example 1

$\min_x f(x) + \|x\|_1$ with $\|\nabla f(x)\|_\infty \leq 1$

$$\tilde{x} = 0 \quad \Rightarrow \quad \langle \nabla f(x), x \rangle + \|x\|_1 \geq 0$$

Example 2

find $x \neq 0$ s.t. $M(x)x = 0$, where M is psd matrix for all x

$$g \equiv 0, \quad \tilde{x} = 0, \quad F(x) = M(x)x \quad \Rightarrow \quad \langle F(x), x \rangle \geq 0$$

$$M(x) = t_1 t_1^T + t_2 t_2^T \quad \text{with} \quad t_1 = A \sin x, t_2 = B \exp x$$

Examples

$$\exists \tilde{x} \in \operatorname{dom} g \quad \text{s.t.} \quad \langle F(x), x - \tilde{x} \rangle + g(x) - g(\tilde{x}) \geq 0 \quad \forall x$$

Example 1

$\min_x f(x) + \|x\|_1$ with $\|\nabla f(x)\|_\infty \leq 1$

$$\tilde{x} = 0 \quad \Rightarrow \quad \langle \nabla f(x), x \rangle + \|x\|_1 \geq 0$$

Example 2

find $x \neq 0$ s.t. $M(x)x = 0$, where M is psd matrix for all x

$$g \equiv 0, \quad \tilde{x} = 0, \quad F(x) = M(x)x \quad \Rightarrow \quad \langle F(x), x \rangle \geq 0$$

$$M(x) = t_1 t_1^T + t_2 t_2^T \quad \text{with} \quad t_1 = A \sin x, t_2 = B \exp x$$

$n = 1000, A, B \in \mathbb{R}^{n \times n}$, $A_{ij}, B_{ij} \sim \mathcal{N}(0,1)$

generate 100 random instances $\Rightarrow$ 100% success rate

Much Ado About

- Lipschitz constant
- Global Lipschitz assumption
- Linesearch

Thanks for attention!