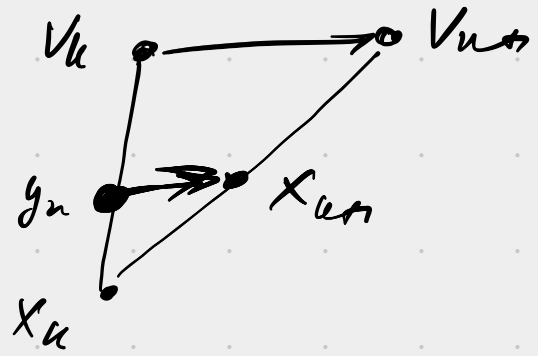
Fast Gradient Method



$$\min_{x \in \mathbb{R}^n} f(x)$$

Init: $x_0 = v_0 \in \mathbb{R}^n$, $A_0 = 0$,

Iteration $k \geq 0$:

1. $y_k = \gamma_k v_k + (1-\gamma_k) x_k$, $\quad \gamma_k \in [0,1]$
2. $v_{k+1} = v_k - a_{k+1} f'(y_k)$, $\quad a_{k+1} > 0$
3. $x_{k+1} = \gamma_k v_{k+1} + (1-\gamma_k) x_k$. $\quad \to f(y_k) - f(x_{k+1}) \geq \frac{1}{2L} \|f'(y_k)\|^2$

$$a_{k+1} = \frac{1}{2L}\left(1 + \sqrt{1 + 4A_k L}\right), \quad A_{k+1} = A_k + a_{k+1}, \quad \gamma_k = \frac{a_{k+1}}{A_{k+1}} \in (0,1].$$

$$x_k = FGM_K(f, x_0)$$

The global convergence:

$$f(x_k) - f^* \leq \frac{2L\|x_0 - x^*\|^2}{k^2}, \quad k \geq 1.$$

Strongly Convex Functions

$$\frac{\mu}{2}\|x_0 - x^*\|^2 \leq f(x_0) - f^*$$

$$f(x_k) - f^* \leq \frac{4L}{\mu K^2}\left(f(x_0) - f^*\right)$$

$$K = \sqrt{\frac{8L}{\mu}} \Rightarrow f(x_k) - f^* \leq \frac{1}{2}\left(f(x_0) - f^*\right)$$

Start $y_0 \in \mathbb{R}^n$.

$$y_{t+1} = FGM_K(f, y_t), \quad t \geq 0$$

Theorem The total complexity of FGM with restarts:

$$\sqrt{\frac{8L}{\mu}} \cdot \log_2 \frac{f(x_0) - f^*}{\varepsilon}.$$

GM: $\frac{L}{\mu} \log_2 \frac{f(x_0)-f^*}{\varepsilon}$

<u>Remarks</u> . We have to know $\frac{L}{\mu}$ for restarts in FGM.

. we can modify FGM $\rightarrow$ it <u>doesn't need restarts</u>. (needs to know $\mu>0$)

. Adaptive search L

Applications. <u>Machine Learning</u>

Generalized Linear Models:

$$f(x) = \frac{1}{m}\sum_{i=1}^{m} \ell(\langle a_i, x\rangle - b_i)$$

x - model parameters , $a_1, \dots, a_m \in \mathbb{R}^n$, $b_1, \dots, b_m \in \mathbb{R}$.

$$\min_{x\in\mathbb{R}^n} \left[f(x) + \psi(x) \right]$$

ψ - "regularizer".

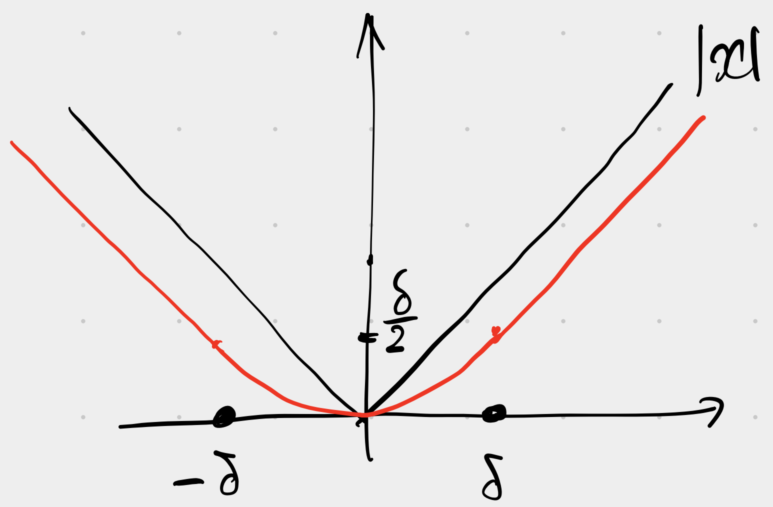
<u>Loss functions</u> $\ell : \mathbb{R} \to \mathbb{R}$ diff convex funct.

. Quadratic loss: $\ell(t) = \frac{1}{2}t^2$, $L_\ell = 1$

. Logistic loss: $\ell(t) = \ln(1+e^t)$, $L_\ell = \frac{1}{4}$.

. Huber loss:

$$\ell(t) = \begin{cases} \frac{1}{2\delta}t^2, & -\delta \le t \le \delta \\ |t| - \frac{\delta}{2}, & \text{otherwise} \end{cases}, \quad \delta > 0$$



$L_\ell = \frac{1}{\delta}$

$f(x) = g(Ax+b)$, $A = \begin{bmatrix} a_1^T \\ \vdots \\ a_m^T \end{bmatrix} \in \mathbb{R}^{m\times n}$, $b = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} \in \mathbb{R}^m$

$g(y) = \frac{1}{m}\sum_{i=1}^{m} \ell(y_i) \Rightarrow \nabla g$ be Lipschitz $L_g = L_\ell$

① Euclidean: $L_f = L_\ell \cdot \|A\|^2$

$v_{k+1} = v_k - a_{k+1} f'(y_k)$ → two matrix-vector products with A.

② Generalized Euclidean: $\|x\| = \langle Bx, x\rangle^{1/2}$, $B = A^TA + \delta I \succ 0$

$m \gg n$

$\underline{L_f = L_\ell}$

$v_{k+1} = v_k - a_{k+1} B^{-1} f'(y_k)$

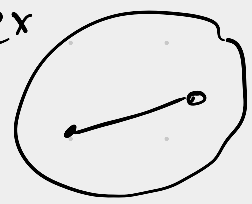
preconditioned step.

<u>Regularization</u>

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

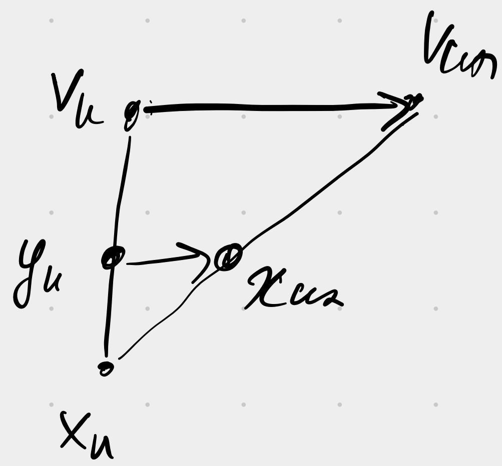
- $\psi(x) = \frac{\mu}{2}\|x\|^2$, $\mu > 0$ ⇒ <u>strongly convex</u>. <u>(weight decay)</u>
- sparsity $\psi(x) = \lambda \|x\|_1$, $\lambda > 0$.
- simple constraints $x \in Q$

Q-convex



$$\psi(x) = \begin{cases} 0, & x \in Q \\ +\infty, & x \notin Q \end{cases} \rightarrow \underline{\text{convex}}$$

$v_{k+1} = \pi_Q(v_k - a_{k+1} f'(y_k))$



Fully Composite Problems

$\min \varphi(x)$, $\varphi(x) = F(x, f_1(x), \dots, f_m(x))$

$f_1, \dots, f_m$ – smooth convex functions : $f_i : \mathbb{R}^n \to \mathbb{R}$

$$f(x) = \begin{bmatrix} f_1(x) \\ \dots \\ f_m(x) \end{bmatrix} : \mathbb{R}^n \to \mathbb{R}^m$$

$$\nabla f(x) = \begin{bmatrix} \nabla f_1(x)^T \\ \dots \\ \nabla f_m(x)^T \end{bmatrix} \in \mathbb{R}^{m \times n} \quad \text{– Jacobian}$$

Lip. grad.: $\|f_i'(x) - f_i'(y)\| \le L_i \|x-y\|$

$$L = \begin{bmatrix} L_1 \\ \vdots \\ L_m \end{bmatrix} \in \mathbb{R}^m$$

- Outer component $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$
  - "simple"

$$\min_x \left[\varphi(x) = F(x, f(x)) \right]$$

- F is jointly convex
- $F(x, \cdot)$ – monotone : $u_1 \le u_2$, $u_1, u_2 \in \mathbb{R}^m \Rightarrow F(x, u_1) \le F(x, u_2)$

$$\Downarrow$$

$\varphi(\cdot)$ is convex

- we assume that F is Lipschitz in u:

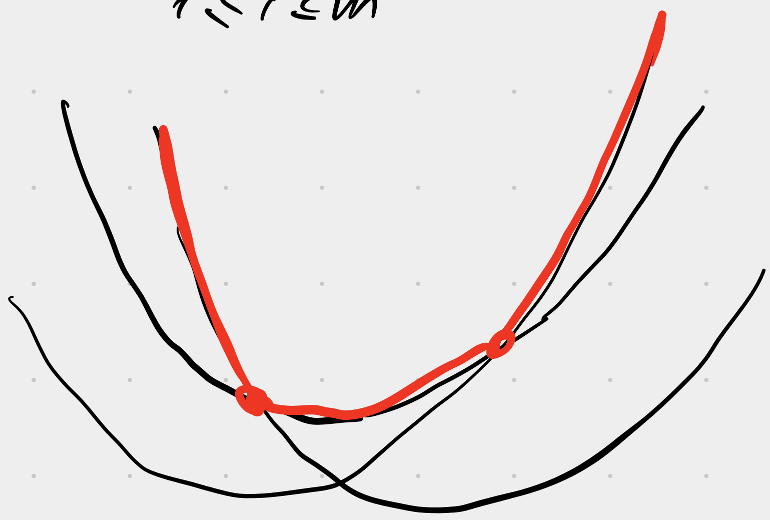
$$|F(x, u_1) - F(x, u_2)| \le M \|u_1 - u_2\| \quad \forall u_1, u_2, x.$$

Examples

- $F(x,u) \equiv u^{(1)}$ then $\varphi(x) = F(x, f(x)) = f_1(x) \to \min_x$
- $F(x,u) \equiv u^{(1)} + \Psi(x)$ then $\varphi(x) = f_1(x) + \Psi(x)$
- Max-type problems $F(x,u) = \max\limits_{1 \le i \le m} u^{(i)}$

$$\varphi(x) = F(x, f(x)) = \max_{1 \le i \le m} f_i(x)$$

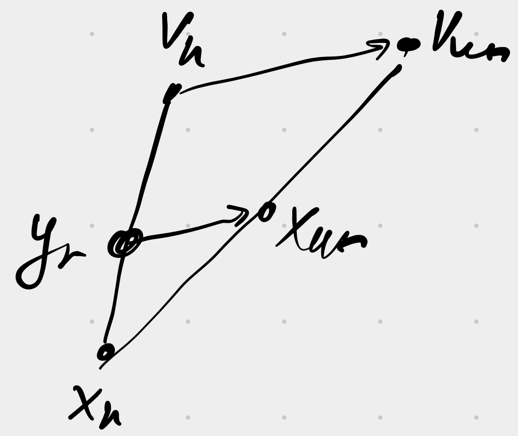
$\varphi(x^*) \le 0.$



Feasibility problem:

$$x^* \in Q = \{x \in \mathbb{R}^n : f_1(x) \le 0, \dots, f_m(x) \le 0\}$$

$\varphi(x) = F(x, f(x))$.



Algorithm: Fully Composite FGM

Init: $x_0 \in \mathrm{dom}\,\psi$. Set $v_0 = x_0$, $A_0 = 0$.

Iterations $k \geq 0$:

1. Choose $a_{k+1} > 0$. $A_{k+1} = A_k + a_{k+1}$, $\gamma_k = \frac{a_{k+1}}{A_{k+1}} \in (0, 1]$.
2. $y_k = \gamma_k v_k + (1-\gamma_k) x_k$, compute $\nabla f(y_k)$
3. Compute v_{k+1} as a solution to:

$$v_{k+1} = \arg\min_x \left[F(x, f(y_k) + \nabla f(y_k)(x - y_k)) + \frac{1}{2a_{k+1}} \|x - v_k\|^2 \right]$$

4. $x_{k+1} = \gamma_k v_{k+1} + (1-\gamma_k) x_k$.

Constr. case $\quad F(x, u) = u^{(1)} + \psi(x)$, $\quad \psi(x) = \begin{cases} 0, & x \in Q \\ +\infty, & x \notin Q \end{cases}$

$$v_{k+1} = \pi_Q(v_k - a_{k+1} f'(y_k))$$

Goal: $\forall k \geq 0$, $\qquad A_k \nearrow$

$$\frac{1}{2}\|x - x_0\|^2 + A_k \varphi(x) \geq \frac{1}{2}\|x - v_k\|^2 + A_k \varphi(x_k), \quad x \in \mathrm{dom}\,\varphi,$$

Assume $k \geq 0$. Consider next iterate $k+1$:

$$\frac{1}{2}\|x - x_0\|^2 + A_{k+1}\varphi(x) = \frac{1}{2}\|x - x_0\|^2 + A_k \varphi(x) + a_{k+1}\varphi(x) \geq$$

$$\geq \frac{1}{2}\|x - v_k\|^2 + A_k \varphi(x_k) + a_{k+1} F(x, f(x))$$

$$\geq \underbrace{\frac{1}{2}\|x - v_k\|^2 + A_k \varphi(x_k) + a_{k+1} F(x, f(y_k) + \nabla f(y_k)(x - y_k))}_{m_k(x) \text{ – strongly convex}} \qquad \min \nearrow (v_{k+1})$$

$$\geq \frac{1}{2}\|x - v_{k+1}\|^2 + m_k(v_{k+1})$$

$$m_k(x) \geq m_k^* + \frac{1}{2}\|x - x_k^*\|^2$$

To show: $m_k(v_{k+1}) \overset{?}{\geq} A_{k+1}\varphi(x_{k+1})$

$$m_k(v_{k+1}) = \frac{1}{2}\|v_{k+1}-v_k\|^2 + \underbrace{A_k F(x_k, f(x_k))} + a_{k+1} F(v_{k+1}, f(y_k) + \nabla f(y_k)(v_{k+1}-y_k))$$

$$\geq \frac{1}{2}\|v_{k+1}-v_k\|^2 + A_k F(x_k, f(y_k) + \nabla f(y_k)(x_k - y_k))$$
$$+ a_{k+1} F(v_{k+1}, f(y_k) + \nabla f(y_k)(v_{k+1}-y_k))$$

$$\geq \underbrace{\frac{1}{2}\|v_{k+1}-v_k\|^2} + A_{k+1} F\Big(\underbrace{\gamma_k v_{k+1} + (1-\gamma_k) x_k}_{x_{k+1}}, f(y_k) + \nabla f(y_k)(x_{k+1}-y_k)\Big)$$

$$\gamma_k = \frac{a_{k+1}}{A_{k+1}}$$

$$\overset{?}{\geq} A_{k+1} F(x_{k+1}, f(x_{k+1})) = A_{k+1}\varphi(x_{k+1})$$

By smoothness:

$$\|f(x_{k+1}) - f(y_k) - f'(y_k)(x_{k+1}-y_k)\| \leq \|L\| \cdot \frac{\|x_{k+1}-y_k\|^2}{2}$$
$$= \|L\| \cdot \underbrace{\frac{\gamma_k^2 \|v_{k+1}-v_k\|^2}{2}}$$

Final choice of a_{k+1}:

$$a_{k+1} = \frac{1}{2\alpha}\left(1 + \sqrt{1 + 4A_k\alpha}\right),$$

$$\alpha = M \cdot \|L\|.$$