Motivation of proximal gradient descent

Friday, March 7, 2025 10:15 AM

$$\underset{x}{\text{Minimize}} \quad \underbrace{f(x)}_{\substack{\text{assume} \\ \text{differentiable} \\ \text{i.e. } \nabla f \text{ is } L\text{-Lipschitz}}} + \underbrace{g(x)}_{\text{assume has easy prox operator}}$$

Algo: prox. grad. descent (aka Forward Backward)

$$x_{k+1} = \text{prox}_{t \cdot g}\left(x_k - t \cdot \nabla f(x_k) \right) \quad \text{for a stepsize } t, \text{ e.g. } t = 1/L$$

Motivation 1: "MM"

our model

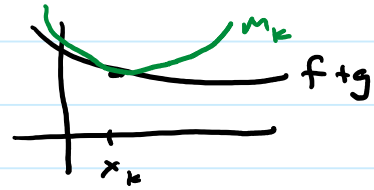
$$x_{k+1} = \underset{x}{\text{argmin}} \; M_k(x) := \underbrace{f(x_k) + \langle \overbrace{\nabla f(x_k)}^{\nabla f_k \text{ for short}}, x - x_k \rangle + \tfrac{L}{2}\|x - x_k\|^2}_{} + g(x)$$

(no approximation to g!)

By Lipschitz property of f, is an upper bound for $f(x)$

... so $M_k(x) \geq f(x) + g(x)$ still an upper bound

and it's exact at x_k.



how to solve for x_{k+1}?

$$M_k(x) = \text{constants} + \underbrace{\langle \nabla f_k, x \rangle + L/2 \|x - x_k\|^2}_{L/2 \| (x - x_k) + 1/L \nabla f_k \|^2} + g(x)$$

$$= \text{const} + L/2 \| x - (\underbrace{x_k - 1/L \nabla f_k}_{x_{k+1/2} \text{ (gradient step)}}) \|^2 + g(x)$$

letting $t = 1/L > 0$ $\qquad x_{k+1/2} = x_k - t \cdot \nabla f_k$

$$x_{k+1} = \underset{x}{\text{argmin}} \; \frac{1}{2t} \| x - x_{k+1/2} \|^2 + g(x)$$

$$= \text{argmin} \; \tfrac{1}{2} \| x - x_{k+1/2} \|^2 + t \cdot g(x)$$

$$= \text{prox}_{tg}(x_{k+1/2}) \quad \checkmark$$

proximal gradient descent (p. 2)

Friday, March 7, 2025 10:15 AM

Motivation 2: Splitting

$$\mathrm{prox}_{tg}(y) = \arg\min_x \tfrac{1}{2}\|x-y\|^2 + t\cdot g(x)$$

Fermat's rule: $0 \in x - y + t\cdot \partial g(x)$

i.e. $y \in (I + \partial g)(x)$

$x = \underbrace{(I+\partial g)^{-1}}(y)$ $\ast$

aka $J_A = (I+A)^{-1}$ is "resolvent" of A

Find $0 \in \partial(f+g)(x)$ Under CQ, find $0 \in \partial f(x) + \partial g(x)$

i.e. $0 \in \nabla f(x) + \partial g(x) \iff 0 \in t\cdot\nabla f(x) + t\cdot \partial g(x)$ $t>0$

$\iff x \in t\,\nabla f(x) + x + t\,\partial g(x)$

i.e.

$\iff x - t\cdot\nabla f(x) \in x + t\,\partial g(x)$

$x_{k+1} = x_k - t\cdot\nabla f(x_k) - t\,\partial g(x_{k+1})$ $\iff (I - t\nabla f)(x) \in (I + t\,\partial g)(x)$

implicit!

$\iff x \overset{\text{if } g\in\Gamma_0(\mathbb{R}^n)}{=} (I + t\,\partial g)^{-1}(I - t\nabla f)(x)$ FIXED PT. EQ'N

So... $x = \mathrm{prox}_{tg}(x - t\cdot\nabla f(x))$

$x = T(x)$ so iterate $x_{k+1} = T(x_k)$

Why "forward backward"?

aka explicit-implicit

ODE $y' = f(t,y)$.

Forward Euler

$y_{k+1} = y_k + h\cdot f(t_k, y_k)$

Backward Euler implicit!

$y_{k+1} = y_k + h\cdot f(t_{k+1}, y_{k+1})$

Stability: apply to

$y' = \lambda\cdot y,\ \lambda<0$

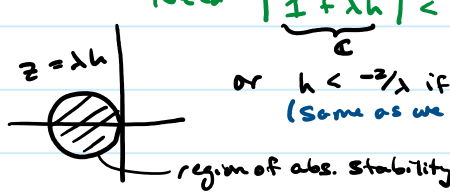
(sol'n: $y(t) = y_0 e^{\lambda t}$)

Forward:

$y_{k+1} = (1+\lambda h)y_k$

$= (1+\lambda h)^k y_0$

Need $|\underbrace{1+\lambda h}_{\mathbb{C}}| < 1$



or $h < -2/\lambda$ if real

(same as we got!)

Backward:

$y_{k+1} = y_k + h\lambda y_{k+1}$

$y_{k+1} = \frac{1}{1-\lambda h} y_k$

$= \left(\frac{1}{1-\lambda h}\right)^k y_0$

Need $\left|\frac{1}{1-\lambda h}\right| < 1$

holds $\forall h>0$!!

i.e. $\mathrm{Re}(\lambda h) < 0$, so

$\lambda<0$, $h\in\mathbb{R}$ ... any h works

unconditionally stable!!

proximal gradient descent (p. 3): linesearch

Friday, March 7, 2025 10:44 AM

For prox. grad. descent, two broad classes of linesearch

(1) Curvilinear (nicer, more costly)

search over $x(t) = \text{prox}_{tg}(x_k - t\nabla f_k)$ recompute prox for every new t

(2) cheaper: $\bar{x} = \text{prox}_{\bar{t}g}(x_k - \bar{t}\nabla f_k)$ for nominal (trial) stepsize $\bar{t}$

then $x(t) = x_k + t\cdot(\bar{x} - x_k)$ starting at $t=1$

Ex: g = indicator of $\mathbb{R}^2_+$