

Gradient descent & Forward Euler

min $f(x)$ where f is smooth.

View as $\frac{dx}{dt} = -\nabla f(x)$

$$\frac{dx}{dt} = -Ax$$

$$x^{(n+1)} = x^{(n)} - \nabla f(x^{(n)}) \Delta t$$

$$x^{(n+1)} = x^{(n)} - Ax^{(n)} \Delta t$$

Converges when Δt suff small.

Thm: If $f: \mathbb{R}^n \rightarrow \mathbb{R}$ convex, smooth

and $\|\nabla f(x) - \nabla f(y)\| \leq L\|x-y\| \quad \forall x,y$ (∇f is L -lipschitz)

then gradient descent w/ step size $h \leq \frac{1}{L}$ converges:

$$f(x^{(n)}) - f(x^*) \leq \frac{\|x^{(0)} - x^*\|^2}{2hn}$$

~~Connection~~

Note: $x \mapsto Ax$ for $A \in \mathbb{R}^{n \times n}$ is Lip w/ const $\|A\|$.

claim: If $h \leq \frac{1}{\|A\|}$

Iterative method for least squares:

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \|Ax - b\|_2^2$$

Solve by grad descent:

$$x^{(n+1)} = x^{(n)} + A^T A x^{(n)} h.$$

Converges when $h \leq \frac{1}{\|A^T A\|} = \frac{1}{\|A\|^2}$.

Activity 0

$$\min \|X\|_1 \text{ s.t. } AX=b$$

Suppose you solve by gradient descent. What happens?

Proximal operator and Backwards Euler

Let $f: \mathbb{R}^n \rightarrow \mathbb{R}$

$$\text{prox}_f(z) = \underset{x}{\text{argmin}} \left(f(x) + \frac{1}{2} \|x - z\|^2 \right)$$

Claim:

$\text{prox}_{hf}(z)$ is a backwards Euler step of size h starting at z

Proof: Let $x^* = \text{prox}_{hf}(z)$

$$\text{so } 0 = h \nabla f(x^*) + (x^* - z).$$

$$\text{Backwards Euler: } \frac{x^* - z}{h} = -\nabla f(x^*) \Rightarrow h \nabla f(x^*) + (x^* - z) = 0.$$

Intuitively

$$\min_x f(x) + \frac{1}{2} \|x - z\|^2$$

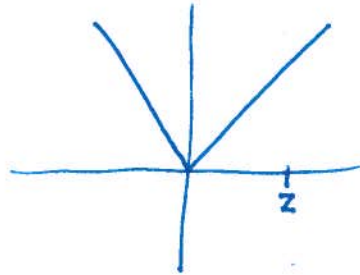
↑
descend
down f

but don't
go too far

Example:

Compute prox for $f(x) = |x|$

$$\min_x |x| + \frac{1}{2}(x-z)^2$$



If z big:

$$\begin{aligned} \min_x x + \frac{1}{2}(x-z)^2 &\Rightarrow 1 + (x-z) = 0 \\ &\Rightarrow x = z-1 \end{aligned}$$

So if $z > 1$, $\text{prox}_{1,1}(z) = z-1$

If z big + negative:

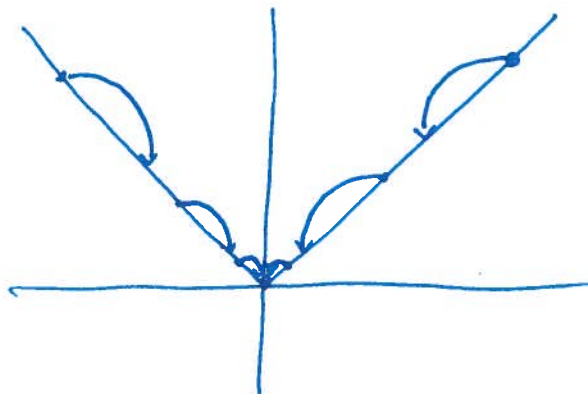
$$\begin{aligned} \min_x -x + \frac{1}{2}(x-z)^2 &\Rightarrow -1 + (x-z) = 0 \\ &\Rightarrow x = z+1 \end{aligned}$$

So if $z < -1$, $\text{prox}_{1,1}(z) = z+1$

If $-1 \leq z \leq 1$, then $\text{prox}_{1,1}(z) = 0$

why: for $0 < z < 1$, $|x| + \frac{1}{2}(x-z)^2 = x + \frac{1}{2}(x-z)^2$ is increasing in x .

Picture:



Soft Thresholding

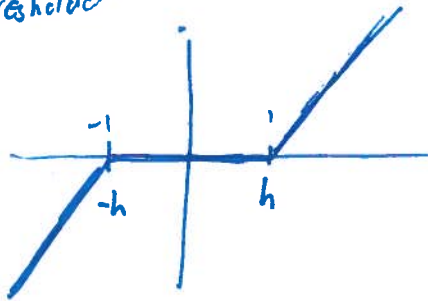
$$z \mapsto \begin{cases} z-1 & \text{if } z \geq 1 \\ 0 & \text{if } |z| \leq 1 \\ z+1 & \text{if } z \leq -1 \end{cases}$$

Proximal operator for ℓ_1 norm. in \mathbb{R}^n

$$\text{prox}_{h\|\cdot\|_1}(z) = \text{Soft-threshold}_h(z) = \left(\left(1 - \frac{h}{|z_i|}\right)_+ z_i \right)_{i=1 \dots n}$$

$\approx z$

Each coef is soft thresholded



Iterative Soft Thresholding for Basis pursuit denoiser

$$\min \|x\|_1 \text{ st } Ax=b$$

write constraint as penalty

$$\min_x \lambda \|x\|_1 + \frac{1}{2} \|Ax-b\|_2^2$$

Idea: Splitting method

- min $\|x\|_1$ term
- min $\|Ax-b\|$ term
- repeat

Forward-backward method

- Do forward (grad descent) step on data misfit term
- Do backward (proximal descent) step on ℓ_1 term

$$X^{(n+\frac{1}{2})} = X^{(n)} + h A^* (b - AX^{(n)}) \quad (\text{grad desc})$$

$$X^{(n+1)} = \text{prox}_{h\lambda\|\cdot\|_1} (X^{(n+\frac{1}{2})})$$

Iterative soft thresholding

$$\text{Converges if } h \in (0, \frac{2}{\|A\|_2^2})$$

Can have adaptive time steps h_n

which can ~~improve~~ improve rate of conv from $O(\frac{1}{n}) \rightarrow O(\frac{1}{n^2})$

Nesterov scheme