Gradient descent & Forward Euler

$\min f(X)$ where f is smooth.

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

$$X^{(n+1)} = X^{(n)} - \nabla f(X^{(n)})\, h \qquad x^{(n+1)} = x^{(n)} - Ax^{(n)} h$$

Converges when Δt suff small.

Thm: If $f: \mathbb{R}^n \to \mathbb{R}$ convex, smooth
and $\|\nabla f(x) - \nabla f(y)\| \le L\|x-y\|$ $\forall x,y$ (∇f is L-Lipschitz)
then gradient descent w/ step size $h \le 1/L$ converges:

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

~~Connection of~~:

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

~~Claim: If $h \le \frac{1}{\ }$~~

Iterativo 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:

$\min \ \|X\|_1$ ~~st Ax=b~~

Suppose you solve by gradient descent. What happens?

Proximal operator and Backwards Euler

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

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

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

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

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

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

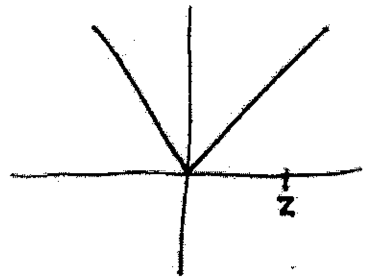
Intuitively

$$\min_x \underset{\substack{\uparrow \\ \text{descend} \\ \text{down } f}}{f(x)} + \underbrace{\tfrac{1}{2}\|x - z\|^2}_{\text{but don't go too far}}$$

Example:

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

$$\min_{x} \ |x| + \tfrac{1}{2}(x-z)^2$$



If z big:

$$\min \ x + \tfrac{1}{2}(x-z)^2 \Rightarrow 1 + (x-z) = 0$$
$$\Rightarrow \ x = z - 1$$

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

If z big + negative:

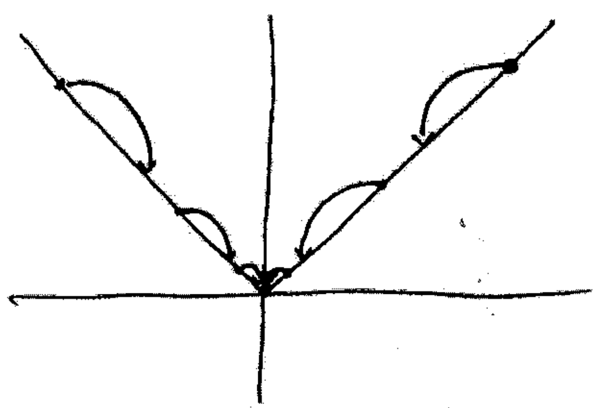
$$\min \ -x + \tfrac{1}{2}(x-z)^2 \Rightarrow -1 + (x-z) = 0$$
$$\Rightarrow \ x = z + 1$$

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

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

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

Picture:



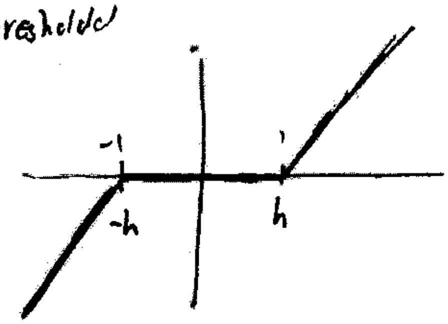
Soft Thresholding

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

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

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

Each coef is soft thresholded



Iterative Soft Thresholding for Basis pursuit denoise

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

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

$$x^{(n+\frac{1}{2})} = x^{(n)} + h A^*\left(b - Ax^{(n)}\right) \qquad \text{(grad desc)}$$

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

Iterative soft thresholding

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

Can have adaptive time steps h_n

which Can ~~more~~ improve rate of conv from $O(\frac{1}{n}) \to O(\frac{1}{n^2})$

Nesterov scheme