Background

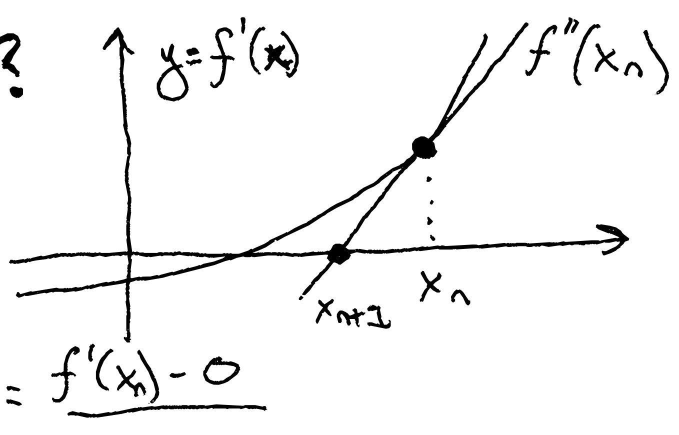
- Sam presented 1st-order alg for cellular sheaves.
- What if we utilized 2nd-order info?
- **Pros:** faster convergence
- **Cons:** challenging to develop...

Newton's Method

Given $f: \mathbb{R}^n \to \mathbb{R}$ C^2 convex,

$x_{n+1} = x_n - (\nabla^2 f(x_n))^{-1} \nabla f(x_n)$

Why?



$$f''(x_n) = \frac{f'(x_n) - 0}{x_n - x_{n+1}}$$

$$\Rightarrow x_{n+1} = x_n - \frac{f'(x)}{f''(x)}$$

Convergence of GD vs. Newton's.

Side Rant about ACT vs. Sheaf Theory

- Newton's is not functorial
- want distributed version which <u>converges</u> not equal to centralized system.

So, let's develop a Newton's for cellular sheaves:

minimize $\sum_i f_i(x_i)$

subject to $L\mathbf{x} = 0$.

We examine "dual Newton's":

$L(\mathbf{x}, \lambda) = \sum_v f_v(x_v) + \lambda^T L \mathbf{x}$ " $\mathbf{x}^T L \lambda$

$q(\lambda) = \inf_{\mathbf{x}} L(\mathbf{x}, \lambda)$

Newton's on q:

$\lambda_{k+1} = \lambda_k + (\nabla^2 q(\lambda_k))^{-1} \nabla q(\lambda_k)$

Let $x(\lambda) = \operatorname{argmin}_x L(x, \lambda)$. Then

$\nabla q(\lambda) = L x(\lambda)$ &

$\nabla^2 q(\lambda) = -L \nabla^2 f(x(\lambda))^{-1} L$.

So $\Delta\lambda_{nt} = (L \nabla^2 f(x(\lambda))^{-1} L)^{-1} L x(\lambda)$.

Let's Break this down:

- $L(\cdot, \lambda)$ is separable across x_v's
  - $L(x, \lambda) = \sum_v L_v(x_v, \lambda)$ where

$$L_v(x_v, \lambda) = f_v(x_v) + x_v^T \sum_{\substack{v \to e \\ u \to e}} F_{v \to e}^T (F_{v \to e} \lambda_v - F_{u \to e} \lambda_u)$$

$\therefore$ $x(\lambda)$ is locally computable!

$\therefore$ $\nabla q(\lambda) = L x(\lambda)$ is " " !

What about the Newton step?

Well, $\nabla^2 f(x(\lambda))$ is blk diag so can be inverted locally.

<u>Challenge</u>: solve the system

$$\star \; \underbrace{(L \nabla^2 f(x(\lambda))^{-1} L)}_{\text{local}} \Delta\lambda_{nt} = \underbrace{L x(\lambda)}_{\text{local}} \; \star$$

Proposed approach: conjugate gradient method.

- CG solves $Ax = b$ for A symmetric psd.
- $\nabla^2 f(x)$ is psd for convex f.
  $\Rightarrow \nabla^2 f(x)^{-1}$ is psd $\Rightarrow L \nabla^2 f(x(\lambda))^{-1} L$ is psd & symmetric!
- CG is a krylov method, only needs to repeatedly apply LMs.
- u & v are conjugate if $u^T A v = 0$.
  - They are "orthogonal wrt A."
- CG is like GD but insists that each search direction is conjugate to previous directions.
- Converges (w/ exact arithmetic) in at most n iters for $A \in \mathbb{R}^{n \times n}$!

<u>Algorithm</u>

$r_0 = b - Ax_0$

$p_0 = r_0$

$k = 0$

do

$\quad \alpha_k = \dfrac{r_k^T r_k}{p_k^T \underbrace{A p_k}}$ ← local

$\quad x_{k+1} = x_k + \alpha_k p_k$ ← local

$\quad r_{k+1} = r_k - \alpha_k A p_k$ ← local

$\quad$ if $r_{k+1} < \varepsilon$, exit

$\quad \beta_k = \dfrac{r_{k+1}^T r_{k+1}}{r_k^T r_k}$

$\quad p_{k+1} = r_{k+1} + \beta p_k$ ← local

$\quad k += 1$

end

return x_{k+1}

$\star$ Only non-local operations are $v^T v$.

- Can compute sums in a totally distributed way using Push-sum.
  - quadratically convergent

Next steps:

- <u>Implementation</u>!
- If that works, prove convergence & publish!