09/17/26: FOLDS seminar, Mateo Diaz

Composite optimization $f = h \circ F$

h convex, F smooth

Ex: nonlinear least squares

Given $b = A(M^*) + \varepsilon$, m linear measurements, recover rank r^* matrix M^*

Ex: X-ray crystallography

Burer-Monteiro approach

$\min_X |A(XX^\top) - b|$

Overparametrize the constraint set of X we're optimizing over

psd rank r can be rewritten as $XX^\top$, and we can search over all $X \in \mathbb{R}^{d \times r}$

But we can also [SJ: ?]

[SJ: To reparametrize what you're saying, basically the choice of reparametrization matters wrt the convergence rate of GD on $h(z)$.]

[SJ: Remind myself what the ...]

[SJ: Think back to what Alex L-N told me about covariance matrix estimation; reparametrizing $\Sigma \mapsto \Sigma^{-1}$ is key.]

Levenberg–Morrison–Marquardt solved this question of preconditioned GD for compositie optinmization that exhibits local convergence...

Solution from '40s: Levenberg–Marquardt algorithm

[SJ: Add algorithm, approximation of Newton, superlinear convergence, see pic, PINN people use this surprisingly (Gomez–Serrano paper)...]

$\min_{x \in \mathbb{R}^d} h(F(x)) = \min_{z \in \mathrm{Im} F} h(z)$

[SJ: See general convergence guarantee, see their "restricted good-conditioning" assumption. Ask for more intuition. The proofs are the same old.]

Exs: nonnegative cone, rank r psd matrices, rank r matrices, symmetric tensors with cp rank r

Topic 2: discretizing and ODE is equivaent to forward discretization or backwards discretization and althougth you can't explictly sovle the backwards discretization (recall the amcs class) you can rewrite it as tan argmin of f plus a squareic term centeref dat previous iterate and do that minimization aka you can do the proximal operator. shervin asks a questions about why backwards discr is better convergence for any gamma stepsize? answer: stability, as evidenced

Also, you can rewrite GD as prox, but not of f, but rather of the linearized f.

Idea: learn a prox operator.

[SJ: Q from D: does error in learning prox operator appear in the other bound, what's the best bound?...]

[SJ: Me: other applications of reformulating GD / forwards / backwards as prox operators throughout gen AI?...]

[SJ: Need to clean up the last few notes.]