# Greedy Newton

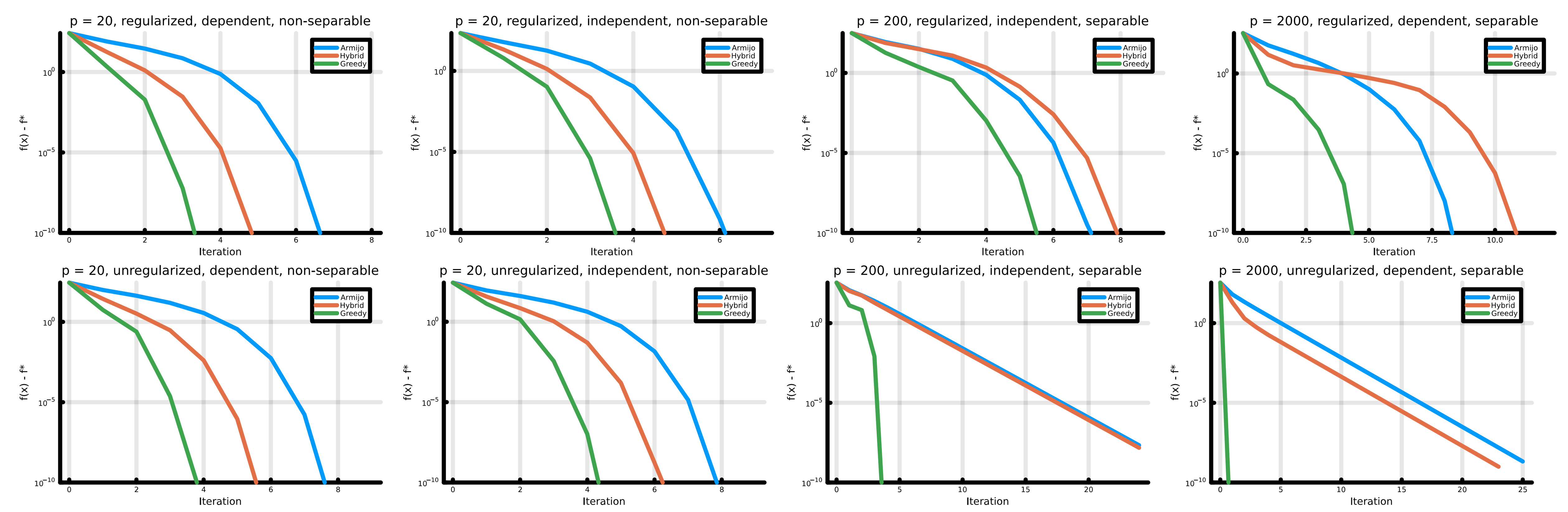
Betty Shea[1] Mark Schmidt[1,2]
[1] University of British Columbia [2] Canada CIFAR AI Chair

Greedy Newton is Newton's Method + **exact** line search

- improves global convergence and preserves local superlinear rate
- experimentally better than (a) Backtracking Newton's Method
  (b) Hybrid-Newton-Gradient Method

*Step sizes could be significantly **larger than 1***


## Greedy Newton vs Hybrid-Newton-Gradient vs Backtracking Newton




## Newton's Method

Iterates in *damped* version

$$x_k = x_k - \alpha_k \nabla^2 f(x_k)^{-1} \nabla f(x_k)$$

uses a **step size** to converge

To find step size, perform **line search** in the direction

$$p_k = -\nabla^2 f(x_k)^{-1} \nabla f(x_k)$$

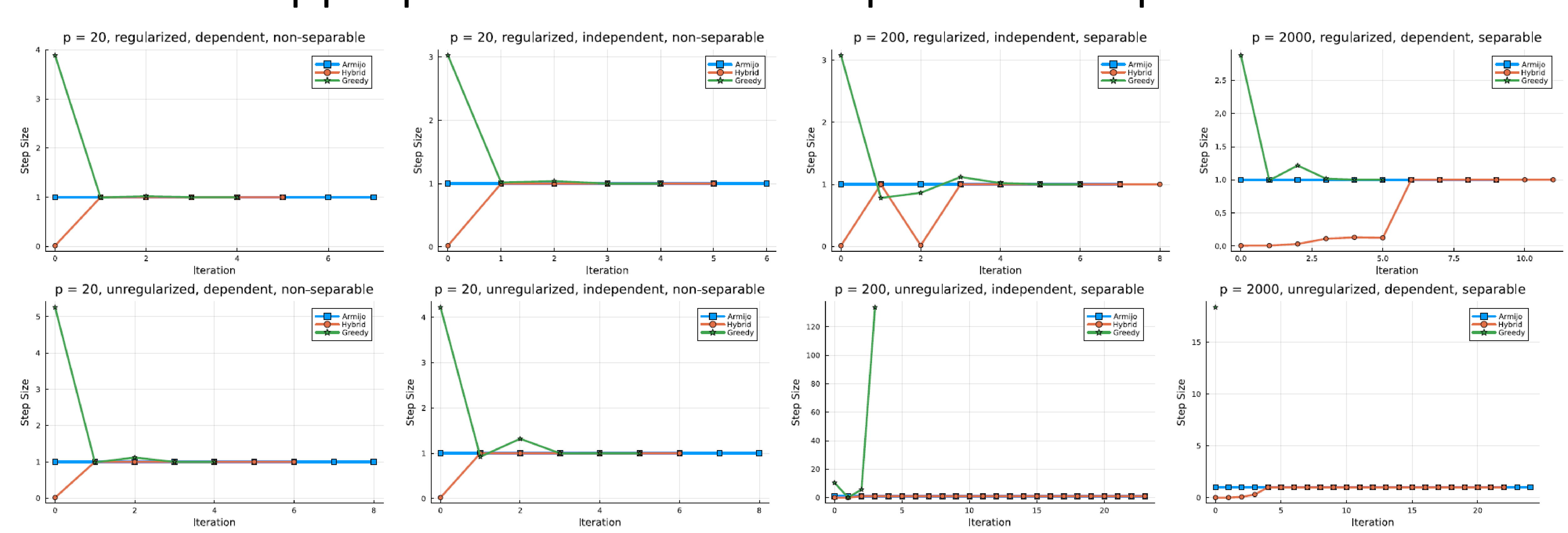
Inexact line search
**Backtracking** Armijo satisfies

$$f(x_k + \alpha_k p_k) \leq f(x_k) + c_1 \alpha_k \nabla f(x_k)^\intercal p_k$$

Exact line search
Finds **greedy** step size with the most 1-step progress

$$\alpha_k \in \arg\min_{\alpha} f(x_k + \alpha p_k)$$

## Greedy Newton adapts to problems by using large step sizes (>>1) at appropriate times in the optimization process




## Future Research

1. At what rate does the Greedy Newton step sizes approach 1?
2. Justification for why Greedy Newton outperforms the theoretically faster hybrid methods?
3. Does a precise line search help with cubic regularized Newton's Method?
4. Can we prove that step sizes larger than 1 improve the global rate?

## Prop. 1 Rate of Greedy Newton

Globally, $$f(x_k) - f(x_*) \leq \left(1 - \frac{\mu^2}{L^2}\right)^k [f(x_0) - f(x_*)]$$

Does as well as backtracking with **any** initial step size, backtracking factor and sufficient decrease condition.

## Prop. 2 Rate of any "as-fast-as-Newton" method

Globally, $$\|x_{k+1} - x_*\| \leq \sqrt{\frac{L}{\mu}} \frac{M}{2\mu} \|x_k - x_*\|^2$$

Local fast convergence when $$\|x_k - x_*\| \leq \sqrt{\frac{\mu}{L}} \frac{2\mu}{M}$$

## Prop. 3 Newton's Method + any step size

Globally,

$$\|x_{k+1} - x_*\| \leq |\alpha_k| \frac{M}{2\mu} \|x_k - x_*\|^2 + |\alpha_k - 1| \frac{L}{\mu} \|x_k - x_*\|$$

Assuming $$|\alpha_k - 1| \leq \|x_k - x_*\|$$

simplifies the global rate to

$$\|x_{k+1} - x_*\| \leq \frac{|\alpha_k| M + 2L}{2\mu} \|x_k - x_*\|^2$$

Local fast convergence when $$\|x_k - x_*\| \leq \frac{2\mu}{|\alpha_k| M + 2L}$$

*Propositions 1-3 assume bounded Hessian* $\mu I \preceq \nabla^2 f(x) \preceq LI$ *Propositions 2 and 3 further assume smooth Hessian* $\|\nabla^2 f(x) - \nabla^2 f(y)\| \leq M\|x - y\|$