# BETTY SHEA[1] MARK SCHMIDT[1,2]
## [1] UNIVERSITY OF BRITISH COLUMBIA  [2] CANADA CIFAR AI CHAIR

2¢ | December 2024

# DON'T BE SO POSITIVE

## NEGATIVE STEP SIZES IN SECOND-ORDER METHODS

**SECOND-ORDER METHODS + NON-CONVEX OBJECTIVES**
- ASCENT DIRECTIONS MAY LEAD TO NON-CONVERGENCE
- DESCENT USUALLY ENFORCED BY DISTORTING CURVATURE INFO
- REDUCES EFFECTIVENESS OF METHOD

**HOW ABOUT TAKING A SUITABLY NEGATIVE STEP INSTEAD?**
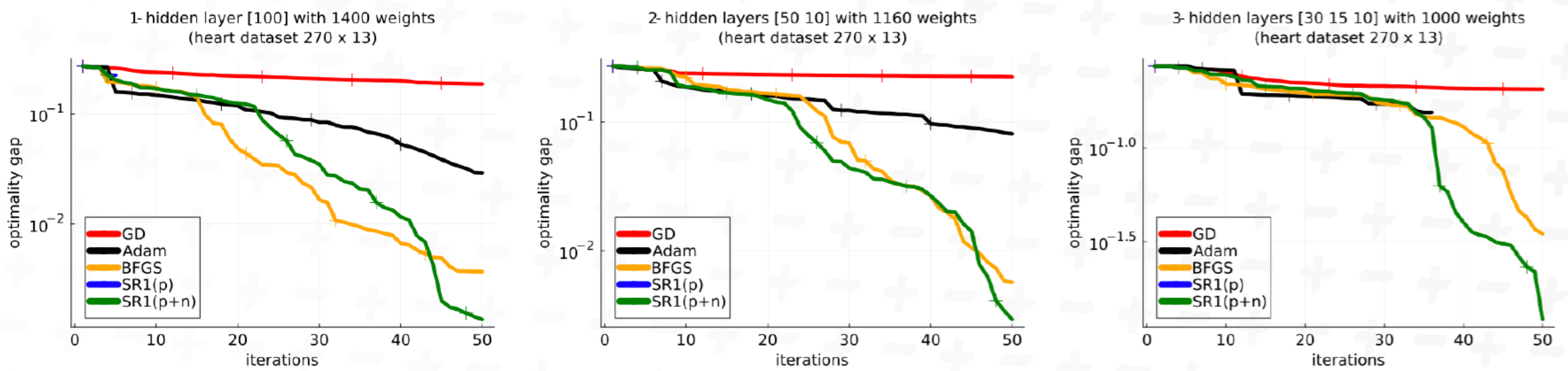


ALLOWING BOTH POSITIVE AND NEGATIVE STEP SIZES (GREEN) MAKES TRAINING CONVERGE FASTER IN NEURAL NETWORKS WITH ONE (LEFT), TWO (MIDDLE) AND THREE (RIGHT) HIDDEN LAYERS.
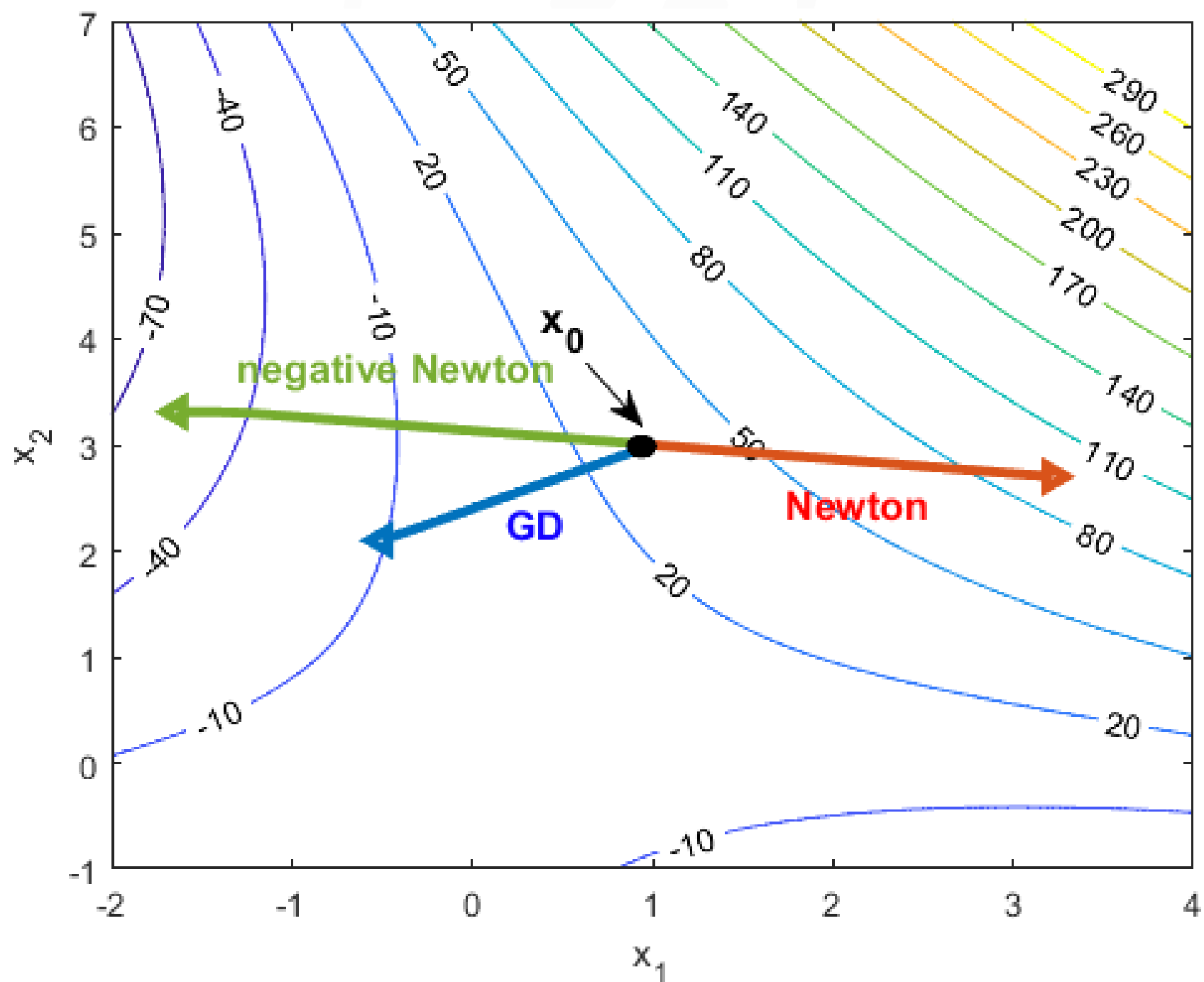


THE NEGATIVE OF NEWTON'S DIRECTION (GREEN) IS A BETTER SEARCH DIRECTION THAN ANY NON-NEGATIVE COMBINATION OF THE GRADIENT DESCENT DIRECTION (BLUE) AND NEWTON'S DIRECTION (RED).

| Dataset | size | GD | Adam | l-BFGS | l-SR1 | l-SR1+damp | l-SR1+Wolfe± |
|---|---|---|---|---|---|---|---|
| a1a | $(1605 \times 119)$ | 0.217 | 0.203 | 0.315 | 0.500 | 0.214 | **0.199** |
| a9a | $(32561 \times 123)$ | 0.220 | 0.212 | 0.332 | 0.330 | 0.216 | **0.211** |
| colon-cancer | $(62 \times 2000)$ | 0.006 | 0.046 | 0.500 | 0.183 | *DNF* | **0.003** |
| gisette | $(6000 \times 5000)$ | 0.139 | 0.171 | 0.395 | 0.395 | *DNF* | **0.050** |
| heart | $(270 \times 13)$ | 0.455 | 0.382 | 0.388 | **0.237** | 0.239 | **0.237** |
| ijcnn1 | $(35000 \times 22)$ | 0.132 | **0.120** | 0.131 | 0.165 | 0.130 | **0.120** |
| ionosphere | $(351 \times 34)$ | 0.201 | **0.118** | 0.201 | 0.500 | 0.199 | 0.170 |
| leukemia | $(38 \times 7129)$ | 0.028 | **0.001** | 0.160 | 0.160 | *DNF* | 0.064 |
| madelon | $(2000 \times 500)$ | 0.500 | 0.500 | 0.490 | 0.396 | **0.382** | 0.396 |
| mushrooms | $(8124 \times 112)$ | 0.040 | 0.018 | 0.166 | 0.500 | 0.019 | **0.009** |
| splice | $(1000 \times 60)$ | 0.270 | 0.260 | 0.485 | 0.500 | 0.262 | **0.195** |
| svmguide3 | $(1243 \times 22)$ | 0.295 | 0.276 | 0.374 | 0.374 | 0.267 | **0.260** |
| w1a | $(2477 \times 300)$ | 0.088 | **0.059** | 0.200 | 0.200 | 0.071 | 0.063 |
| w8a | $(49749 \times 300)$ | 0.096 | **0.066** | 0.193 | 0.193 | 0.079 | 0.071 |

EXPERIMENTS USING VARIOUS LIBSVM DATASETS. ALLOWING NEGATIVE STEP SIZES WORK BETTER THAN HESSIAN MODIFICATIONS FOR METHODS THAT PRODUCE BOTH ASCENT AND DESCENT SEARCH DIRECTIONS.

**SCAN ME!**