Today

Finish up Optimization

Adam

Announcements

HW 7 due tomorrow

Exam grades ASAP

Algorithms for
- C^2-continuous
- Unconstrained

Optimization

Name	Step size control	Steps to solve Rosenbrock	Notes
Newton's	B: Backtracking (Armijo or Wolfe - more complex)	~20	Quadratic Approximation + Very good direction + initial step size + Quadratic convergence near optimum - Requires Hessian! - Requires memory to store Hessian
Quasi-Newton (BFGS)	B	~30	Estimates Hessian as you go with gradients + Good direction + Only requires gradients - Complex to implement - Requires enough memory to store Hessian
Adam	Adaptive based on average square gradient	~2000 (fewer steps)	Momentum and squared gradient step size O Decent direction and step size + Tolerant of noisy gradient estimates + No line search needed (fewer function evals/gradient eval)
Conjugate Gradient	B	~2000	Use history of past gradients to get better search direction O Decent direction + low memory + only requires gradient
Gradient (steepest) descent	B	~10000	- Very bad direction + Easy to find direction (just gradient)

How did Danny score 10.5M on Lab 1

$A\vec{x} = \vec{b}$

minimize $\frac{1}{2}\vec{x}^T A\vec{x} - b\vec{x} = f(\vec{x})$

$\nabla f(\vec{x}) = A\vec{x} - \vec{b} = 0$

FONC is satisfied when $A\vec{x} = b$

Use conjugate gradient with $\nabla f(\vec{x})$

Adam

input : starting point $\vec{x}_0$, step factor α, decay rates $\gamma_v, \gamma_s \in [0,1)$, small constant ϵ, tolerance E_s

output: $\vec{x}^*$, estimate of the minimizer, and $f(\vec{x}^*)$

// defaults from the original paper: $\alpha = 0.001$, $\gamma_v = 0.9$, $\gamma_s = 0.999$, $\epsilon = 10^{-8}$

// all vector products, square roots, and divisions below are elementwise

$\vec{v}_0 \leftarrow \vec{0}, \;\; \vec{s}_0 \leftarrow \vec{0}, \;\; k \leftarrow 0$

while $\|\nabla f(\vec{x}_k)\| \;\; > E_s$ **do**

velocity

acceleration

integration

Like Euler

$\vec{g}_k \leftarrow \nabla f(\vec{x}_k)$

$\vec{v}_{k+1} \leftarrow \gamma_v \vec{v}_k + (1-\gamma_v)\, \vec{g}_k$ // biased decaying momentum

$\vec{s}_{k+1} \leftarrow \gamma_s \vec{s}_k + (1-\gamma_s)\, (\vec{g}_k \odot \vec{g}_k)$ // biased decaying sq. gradient

$\hat{\vec{v}}_{k+1} \leftarrow \vec{v}_{k+1} / (1-\gamma_v^{k+1})$ // corrected decaying momentum

controls step sizes in all direction

$\hat{\vec{s}}_{k+1} \leftarrow \vec{s}_{k+1} / (1-\gamma_s^{k+1})$ // corrected decaying sq. gradient

$\vec{x}_{k+1} \leftarrow \vec{x}_k - \dfrac{\alpha}{\epsilon + \sqrt{\hat{\vec{s}}_{k+1}}} \odot \hat{\vec{v}}_{k+1}$ // next iterate

$k \leftarrow k+1$

end

return $\vec{x}^* = \vec{x}_k, \;\; f(\vec{x}^*)$

In the lab $\beta_1 = \gamma_v$

$\beta_2 = \gamma_s$