Lecture 26 Backtracking Line Search

Note by Samion Suwito on 4/24/25

Announcements

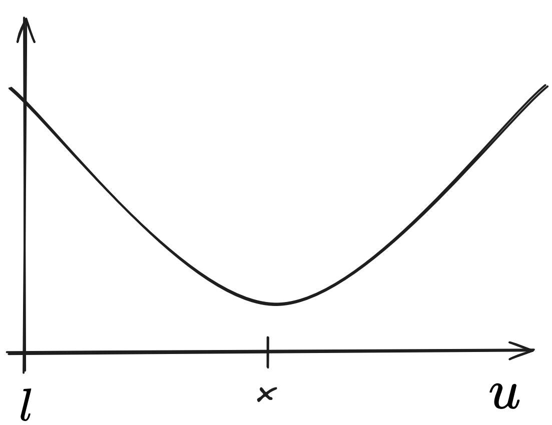
- Previously: 1 Automatic HW Drop + 1 Drop from Survey
  - Switched to 2 Automatic HW Drop + 1 Drop from Survey
- Previously: 50% Midterm Clobber
  - Switched to full midterm clobber
- 2 Review sessions during RRR week at normal time
- Around 6 questions on Final Exam
- Discussions end next Wednesday

Last Time

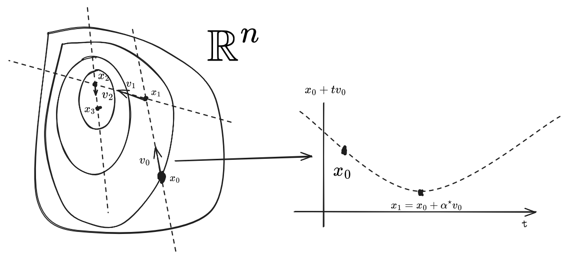
Assuming suitable constraint qualification (E.g. Slater),
We can solve **constrained optimisation problems** by solving a sequence of **unconstrained optimisation problems**.
Ex. Important Phase I/II by running Interior Point Method (IPM) for decreasing parameter t

So we will still need to understand how to solve unconstrained problems. Many possible approaches (gradient descent, Newton's Method, Gauss-Seidel, Stochastic GD,…)

1-dim optimisation: Can solve by bisection **Lecture 25 Algorithms 2 > Unconstrained 1D Minimisation**:

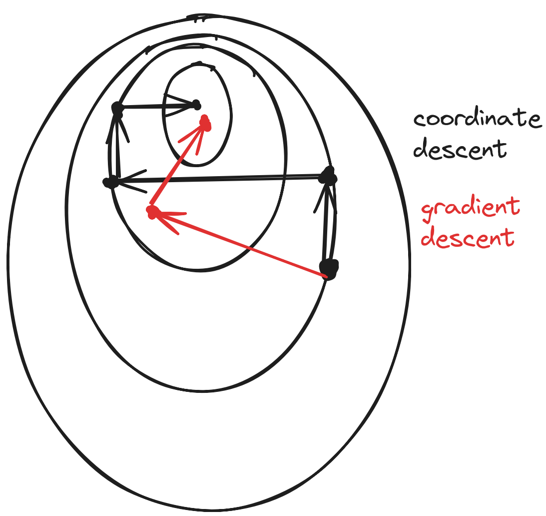


Extension to n dimensions



Given starting point x_0 and direction v_0, solve $\alpha^\star = \arg\min_{t\in\mathbb{R}} f_0(x_0 + tv_0)$. Update $x_0 \leftarrow x_0 + \alpha^\star v_0$ Choose new descent direction v_0, repeat.

Question: How to choose descent direction v_0 at each step
Example 1: Coordinate Descent: let v_0 loop among coordinate directions (in $\mathbb{R}^n$ it's $\{e_1, \dots, e_n\}$)
Example 2: Gradient Descent: $v_0 = -\nabla f_0(x_0)$
Example 3: Newton's Method: choose v_0 in the direction you would travel if objective were quadratic.

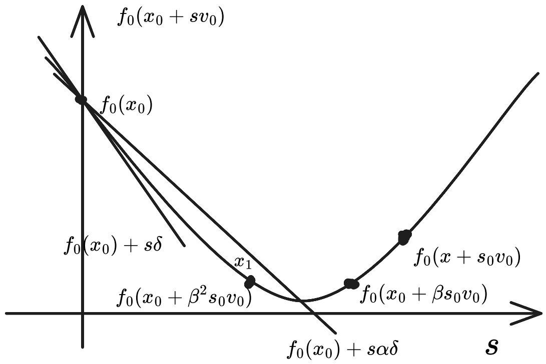


At each step solving for optimal step size $\alpha^\star$ is overkill. Don't need to solve exactly it suffices to solve approximately:

Backtracking line search

Given point x_0, direction v_0, parameters $\alpha, \beta \in (0,1)$ and $s_0 =$ initial step size, (typically $s_0 = 1$)

1. Set $s = s_0, \delta = \nabla f_0(x_0)^\top v_0$
2. If $f_0(x_0 + sv_0) \leq f_0(x_0) + s\alpha\delta$, then $x_0 \leftarrow x_0 + sv_0$, otherwise $s \leftarrow \beta s$ then repeat step 2.



Bisection is really trying to solve the equation $f'(x) = 0$ where $x \in \mathbb{R}$. For a convex function, $f'' \geq 0 \implies f'$ is nondecreasing. $f'(x)$