16-745 Optimal Control Lecture 4

Reid Graves

January 26, 2025

1 Last Time

- Root Finding
- Newton's Method
- Minimization
- Regularization

2 Today

- Line Search (fix overshooting problem)
- Constrained Minimization

3 Line Search

- Often Δx step from Newton overshoots the minimum
- To fix this, check $f(x + \Delta x)$ and "backtrack" untill we get a "good" reduction
- Many strategies
- A simple and effective strategy is "Armijo Rule"

$$\alpha = 1$$

while $f(x + \alpha\Delta x) > f(x) + b\alpha\nabla f(x)^T\Delta x$

b is tolerance, whole addition to $f(x)$ is expected reduction from linearization

$\alpha \leftarrow c\alpha$ c is a scalar < 1

end

- Intuition
  - Make sure step agrees with linearization within some tolerance b
- Typical values:

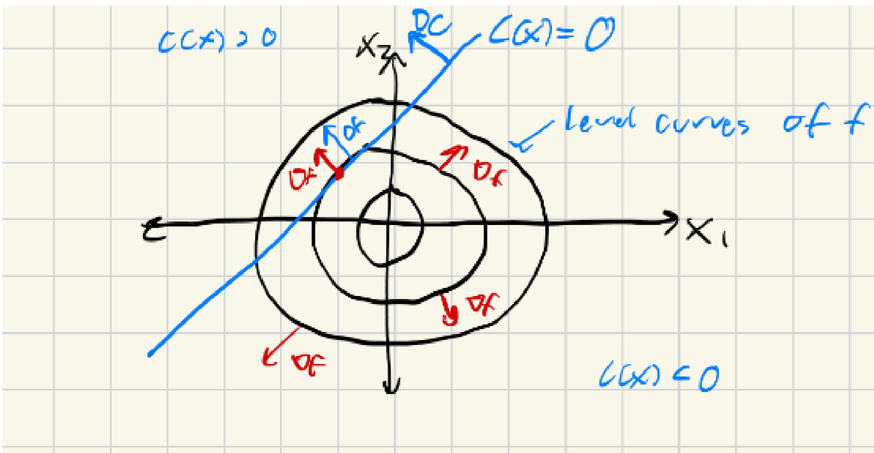
$$c = \frac{1}{2}, b = 10^{-4} \text{ to } 0.1$$

- Take away:
  - Newton with simple and cheap modifications called "globalization strategies" is extremely effective at finding local optima

4 Equality constraints

$$\min_x f(x) \leftarrow f(x) : \mathbb{R}^n \to \mathbb{R}$$
$$\text{s.t. } c(x) = 0 \leftarrow c(x) : \mathbb{R}^n \to \mathbb{R}^m$$

- First-Order necessary conditions
  1. Need $\nabla f(x) = 0$ in **free** directions
  2. Need $c(x) = 0$



$$f(x) : \mathbb{R}^2 \to \mathbb{R}$$
$$c(x) : \mathbb{R}^2 \to \mathbb{R}$$

- Any nonzero component of ∇f must be normal to the constraint at an optimum- equivalently ∇f must be parallel to ∇c

$$\Rightarrow \nabla f + \lambda \nabla c = 0$$

λ is Lagrange multiplier/"dual variable"

- In general:

$$\frac{\partial f}{\partial x} + \lambda^T \frac{\partial c}{\partial x} = 0, \quad \lambda \in \mathbb{R}^m$$

- Based on this gradient condition, we define:

$$L(x, \lambda) = f(x) + \lambda^T c(x)$$

L is lagrangian

- Such that:

$$\nabla_x L(x,\lambda) = \nabla f + (\frac{\partial c}{\partial x}^T \lambda = 0$$
$$\nabla_\lambda L(x,\lambda) = C(x) = 0$$

- We can sove this with Newton:

$$\nabla_x L(x+\Delta x, +\lambda+\Delta\lambda) \approx \nabla_x L(x,\lambda) + \frac{\partial^2 L}{\partial x^2}\Delta x + \frac{\partial^2 L}{\partial x \partial \lambda}\Delta\lambda = 0$$
$$\nabla_\lambda L(x+\Delta x, \lambda+\Delta\lambda) \approx c(x + + \frac{\partial c}{\partial x} = 0$$

$$\begin{bmatrix} \frac{\partial^2 L}{\partial x^2} & \left(\frac{\partial c}{\partial x}\right)^T \\ \frac{\partial c}{\partial x} & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta \lambda \end{bmatrix} = \begin{bmatrix} -\nabla_x L(x,\lambda) \\ -c(x) \end{bmatrix}$$

This is “KKT system”

- Gauss-Newton Method:

$$\frac{\partial^2 L}{\partial x^2} = \nabla^2 f + \frac{\partial}{\partial x}\left[\left(\frac{\partial c}{\partial x}\right)^T \lambda\right]$$

Right additive term is expensive to compute

- We often drop the 2nd “Constraint curvature” term
- Called “Gauss Newton”
- Slightly slower convergence than full Newton method (more iterations) but iterations are cheaper
- Often wins in wall-clock time

4.1 Example

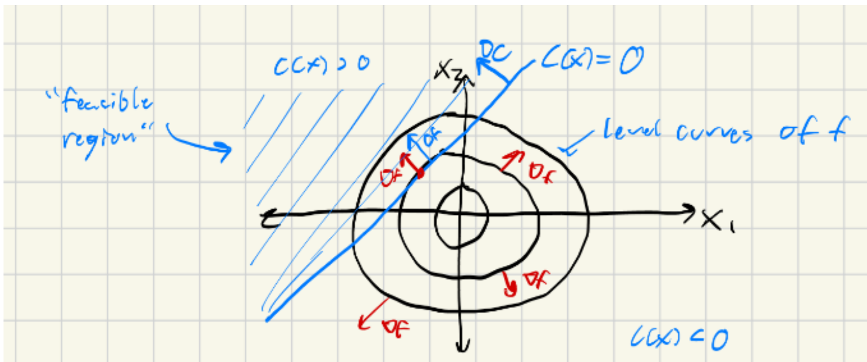
- start $[-1,-1]$, $[-3,2]$ Full newton got stuck on $[-3,2]$, but Gauss-Newton doesn't
- Take Aways:
  - May still need to regularize $\frac{\partial^2 L}{\partial x^2}$ even if $\nabla^2 f > 0$
  - Gauss-Newton is often used in practice

5 Inequality Constraints

$$\min_x f(x), \text{ s.t. } c(x) > 0$$

- We'll just look at inequalities for now
- Just combine with previous methods to handle both kinds of constraints

- First-Order Necessary Conditions:
  1. $\nabla f = 0$ in the **free directions**
  2. $c(x) \geq 0$



- KKT conditions:

$$\begin{aligned} \nabla f - \left(\frac{\partial c}{\partial x}\right)^T \lambda = 0 \quad &\leftarrow \text{Stationarity} \\ c(x) \geq 0 \quad &\leftarrow \text{Primal feasibility} \\ \lambda \geq \quad &\leftarrow \text{Dual feasibility} \\ \lambda \cdot c(x) = \lambda^T c(x) = 0 \quad &\leftarrow \text{Complementarity} \end{aligned}$$