Last Time:

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

Today:

- Line Search
- Constrained Minimization

---

* Line Search

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

$\alpha = 1$

$$\text{while } f(x+\alpha\Delta x) > f(x) + \overset{\text{tolerance}}{b}\,\underbrace{\alpha \nabla f(x)^T \Delta x}_{\text{expected reduction from linearization}}$$

$$\alpha \leftarrow \overset{}{c}\,\alpha$$

scalar < 1

end

* Intuition:

  - Make sure step agrees with linearization within some tolerance b

* Typical values

$$c = 1/2 \ , \quad b = 10^{-4} \sim 0.1$$

* Take Away

  - Newton with simple + cheap modifications ("globalization strategies") is extremely effective at finding local optima.

---

* Equality Constraints

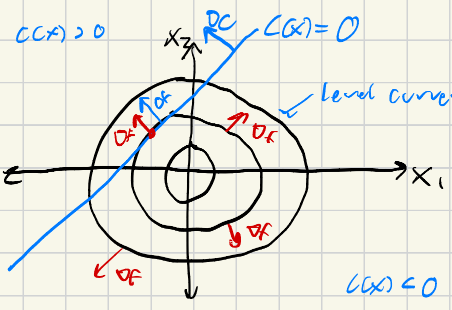
$$\min_x \ f(x) \quad \leftarrow \quad f(x): \mathbb{R}^n \rightarrow \mathbb{R}$$

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

  - First-Order Necessary Conditions

    1) Need $\nabla f(x) = 0$ in <u>free</u> directions

    2) Need $c(x) = 0$



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

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

* Any nonzero component of ∇f must be normal to the constraint at an optimum. Equivalently ∇f must be parallel to ∇c

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

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:

$$\underbrace{L(x,\lambda)}_{\text{Lagrangian}} = f(x) + \lambda^T c(x)$$

- Such that

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

$$\nabla_\lambda L(x,\lambda) = c(x) = 0$$

- We can solve this with Newton:

$$\nabla_x L(x+\Delta x, \lambda+\Delta\lambda) \approx \nabla_x L(x,\lambda) + \frac{\partial^2 L}{\partial x^2}\Delta x + \underbrace{\frac{\partial^2 L}{\partial x \partial \lambda}}_{\left(\frac{\partial c}{\partial x}\right)^T}\Delta\lambda = 0$$

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

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

* Gauss-Newton Method:

$$\frac{\partial^2 L}{\partial x^2} = \nabla^2 f + \underbrace{\frac{\partial}{\partial x}\left[\left(\frac{\partial c}{\partial x}\right)^T \lambda\right]}_{\text{This term is expensive to compute}}$$

- We often drop the 2nd "constraint curvature" term
- Called "Gauss Newton"
- Slightly slower convergence full Newton (more iterations) but iterations are cheaper

$\Rightarrow$ often wins in wall-clock time

* Example:

- start $[-1, -1]$, $\underbrace{[-3, 2]}$ full Newton gets stuck Gauss-Newton doesn't

* Take Aways:

- May still need regularize $\frac{\partial^2 L}{\partial x^2}$ even if $\nabla^2 f > 0$
- Gauss-Newton is often used in practice

* Inequality Constraints:

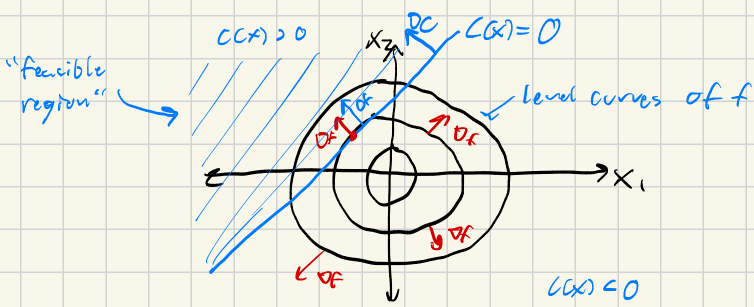
$$\min_x f(x)$$
$$\text{s.t. } c(x) \geq 0$$

- 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

$$\nabla f - \left(\frac{\partial c}{\partial x}\right)^T \lambda = 0 \leftarrow \text{"stationarity"}$$

$$c(x) \geq 0 \leftarrow \text{"primal feasibility"}$$

$$\lambda \geq 0 \leftarrow \text{"dual feasibility"}$$

$$\lambda \circ c(x) = \lambda^T c(x) = 0 \leftarrow \text{"complementarity"}$$