Last Time:

- Root Finding
- Newton's Method
- Minimizing
- 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"

$\alpha = 1$ ← "step length"

tolerance

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

$\quad \alpha \leftarrow c\alpha$

scalar < 1

end

* Intuition:

  - Make sure step agrees with linearization within some tolerance σ

* Typical values

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

* Take Away

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

---

* Equality Constraints

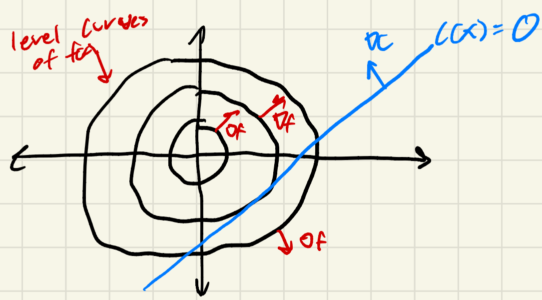
$$\min_x \ f(x) \quad \leftarrow f(x) : \mathbb{R}^n \to \mathbb{R}$$

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

- First-Order Necessary Conditions:

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

  2) Need $c(x) = 0$



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

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

* Any non-zero component of ∇f must be normal to the constraint surface/manifold

$$\Rightarrow \quad \nabla f + \underbrace{\lambda}_{\text{Lagrange multiplier / "dual variable"}} \nabla c = 0$$

* In general:

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

- 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:

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

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

$$\Rightarrow \frac{\partial c}{\partial x}\Delta x = -c(x)$$

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

* 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 than full Newton (more iterations) but iterations are cheaper

$\Rightarrow$ often wins in wall-clock time

* Example:

-start at $[-1,-1]$, $[-3,2]$

full Newton gets stuck
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

---

* Inequality Constraints

$$\min_x \; f(x)$$
$$\text{s.t.} \quad c(x) \leq 0$$

- 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 <u>free</u> directions

2) $c(x) \leq 0$

$$
\text{KKT conditions}\;\begin{cases}
\nabla f + \left(\frac{\partial c}{\partial x}\right)^T \lambda = 0 & \leftarrow \text{“stationarity”} \\
c(x) \le 0 & \leftarrow \text{“primal feasibility”} \\
\lambda \ge 0 & \leftarrow \text{“dual feasibility”} \\
\lambda \circ c(x) = \lambda^T c(x) = 0 & \leftarrow \text{“complementarity”}
\end{cases}
$$

* Intuition:

- If constraint is "active" $\underbrace{(c(x) = 0) \Rightarrow \lambda > 0}_{\text{same as equality case}}$

- If constraint is "inactive" $\underbrace{(c(x) < 0) \Rightarrow \lambda = 0}_{\text{same as unconstrained case}}$

- Complementarity encodes "on/off" switching