

Last Time:

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

Today:

- Line Search
- Constrained Minimization

* Line Search

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

$$\alpha = 1 \leftarrow \text{"step length"}$$

tolerance

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

expected reduction
from linearization

$$x \leftarrow c\bar{x}$$

$$c \text{ scalar} < 1$$

end

* Intuition:

- Make sure step agrees with linearization within some tolerance b

* Typical Values

$$C = \gamma_2, \quad b = 10^{-4} - 0.1$$

* Take Away

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

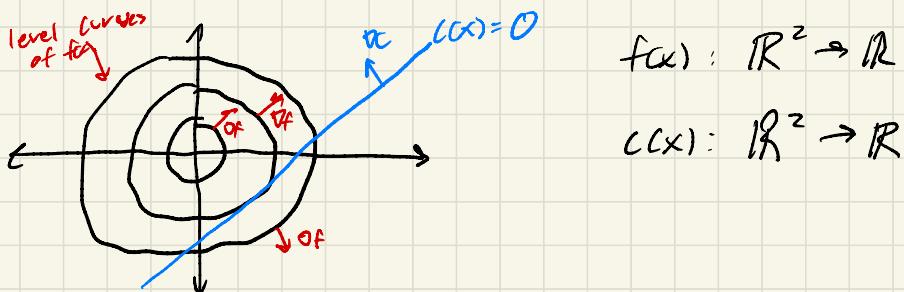
* Equality Constraints

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

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

- First-Order Necessary Conditions:

- 1) Need $\nabla f(x) = 0$ in free direction
- 2) Need $C(x) = 0$



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

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

λ
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:

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

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

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

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

"KKT System"

* Gauss - Newton Method:

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

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 $[-1, -1]$, $\underbrace{[-3, 2]}$

Full Newton gets stuck
 Gauss-Newton doesn't

* Take Aways:

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

* Inequality Constraints

$$\min_x f(x)$$

$$\text{s.t. } C(x) \leq 0$$

- 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 free directions
- 2) $C(x) \leq 0$

KKT conditions

$$\left\{ \begin{array}{l} \nabla f + \left(\frac{\partial L}{\partial x}\right)^T \lambda = 0 \quad \leftarrow \text{"stationarity"} \\ C(x) \leq 0 \quad \leftarrow \text{"primal feasibility"} \\ \lambda \geq 0 \quad \leftarrow \text{"dual feasibility"} \\ \lambda^T C(x) = \lambda^T C(x) = 0 \quad \leftarrow \text{"complementarity"} \end{array} \right.$$

* Intuition:

- If constraint is "active" ($C(x) = 0$) $\underbrace{\Rightarrow \lambda > 0}_{\text{same as equality case}}$
- If constraint is "inactive" ($C(x) < 0$) $\underbrace{\Rightarrow \lambda = 0}_{\text{same as unconstrained case}}$
- Complementarity encodes "on/off" switching