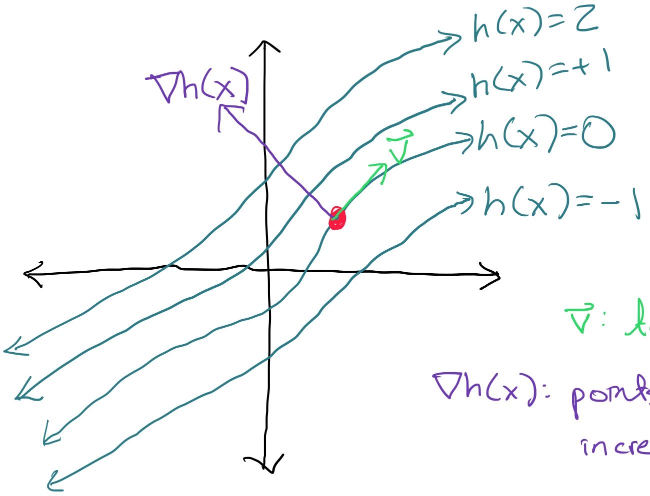
Lecture 5 2025-09-09

Last time: rank-nullity theorem, least squares

Today: least norm, equality-constrained Newton method

Revisit necessary conditions of optimality:



$\vec{v}$: tangent to level surface

$\nabla h(x)$: points in direction of greatest increase

1. steps along $\vec{v}$ remain on $h(x)=0$
2. steps along $\pm\nabla h(x)$ violate $h(x)=0$

$$\nabla f(x^{(k)}) = \mu \vec{v} + \lambda \nabla h(x^{(k)})$$

$$\nabla f(x^{(k)}) = m\vec{v} + \lambda \nabla h(x^{(k)})$$

(since $\vec{v}$ and $\nabla h(x^{(k)})$ are orthogonal)

→ can only move along component of $\nabla f(x^{k})$ parallel to $\vec{v}$

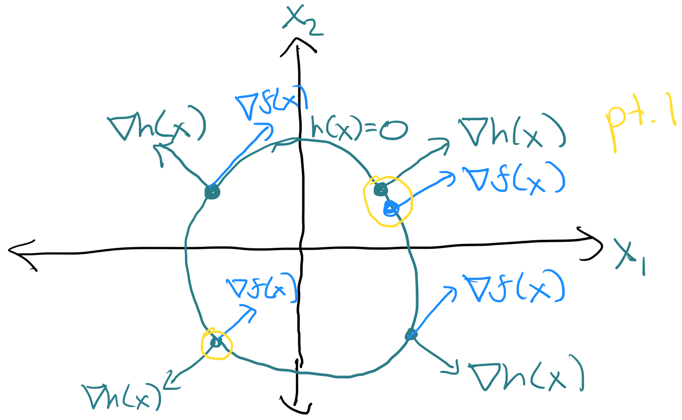
→ stopping condition: at x^*, there cannot be a step we take in $\vec{v}$ direction

→ at x^* : $\nabla f(x^*) = \lambda \nabla h(x^*)$

e.g.

$$\min_{x} \quad x_1 + x_2$$

subj. to: $x_1^2 + x_2^2 = 4$ → $h(x) = x_1^2 + x_2^2 - 4 = 0$



$\nabla f(x) = (+1, +1)$ $\qquad$ $\nabla h(x) = (2x_1, 2x_2)$

Note there are two points satisfying N.C.O:

pt 1: $\nabla f(x) = +2\, \nabla h(x)$

pt 2: $\nabla f(x) = -2\, \nabla h(x)$

Least norm problem:

$$\min_x \ \frac{1}{2} x^T x$$

subj. to: $Ax = b$ $\qquad$ where $A \in \mathbb{R}^{m \times n}$ and full

$(m \leq n)$

rank

defined: $\mathcal{L}(x,y) = f(x) + y^T(Ax-b)$

$y \in \mathbb{R}^m$ is Lagrange multiplier

$\nabla_x \mathcal{L}(x^*, y^*) = 0$

$\mathcal{L}(x,y) = f(x) + y^T h(x) = \frac{1}{2} x^T x + y^T(Ax-b)$

$\nabla_x \mathcal{L} = x + A^T y = 0 \rightarrow x = -A^T y$

$\nabla_y \mathcal{L} = 0 = Ax - b = -AA^T y - b = 0$

$\rightarrow y = -(AA^T)^{-1} b$

$\rightarrow$ plug back: $x = -A^T y = A^T(AA^T)^{-1} b$

least norm solution: $x^* = A^T(AA^T)^{-1} b$

Generalize: $\min_x f(x)$ $f \in \mathcal{C}^2$ (Hessian exists)

subj. to $Ax = b$

and use Newton method to solve for x^*

solution approach: find x^* such that N.C.O. satisfied

$$\mathcal{L}(x) = f(x) + y^T(Ax-b)$$

N.C.O.: $\nabla_x \mathcal{L}(x^*) = 0 = \nabla_x f(x) + A^T y$

$\nabla_y \mathcal{L}(x^*) = 0 = Ax - b$

define residual function:

$$r(x,y) = \begin{pmatrix} \nabla_x \mathcal{L}(x,y) \\ \nabla_y \mathcal{L}(x,y) \end{pmatrix}$$

solving for $x^* \Longleftrightarrow$ solving for $r(x,y) = 0$

solving

solving for r(x,y

given reference $(\bar{x},\bar{y})$, take Taylor expansion of $r(x,y)$

$$\rightarrow r(x,y) \approx r(\bar{x},\bar{y}) + \underbrace{\left.\frac{\partial r}{\partial x}\right|_{(\bar{x},\bar{y})}}_{\substack{\text{Jacobian} \\ \text{w.r.t. } x}} \Delta x + \underbrace{\left.\frac{\partial r}{\partial y}\right|_{(\bar{x},\bar{y})}}_{\substack{\text{Jacobian} \\ \text{w.r.t. } y}} \Delta y := 0$$

Jacobian w.r.t. x:

$$\frac{\partial r}{\partial x} = \left.\begin{pmatrix} \frac{\partial}{\partial x} \nabla_x \mathcal{L}(x,y) \\ \frac{\partial}{\partial x} \nabla_y \mathcal{L}(x,y) \end{pmatrix}\right|_{(\bar{x},\bar{y})}$$

$$= \begin{pmatrix} \frac{\partial}{\partial x} \left(\nabla f(x) + A^T y\right) \\ \frac{\partial}{\partial x} (Ax - b) \end{pmatrix} = \begin{pmatrix} \nabla_{xx} f(\bar{x}) \\ A \end{pmatrix}$$

Jacobian w.r.t. y:

$$\frac{\partial r}{\partial y} = \left.\begin{pmatrix} \frac{\partial}{\partial y} \nabla_x \mathcal{L}(x,y) \\ \frac{\partial}{\partial y} \nabla_y \mathcal{L}(x,y) \end{pmatrix}\right|_{(\bar{x},\bar{y})}$$

$$= \begin{pmatrix} \frac{\partial}{\partial y} \left(\nabla f(x) + A^T y\right) \\ \frac{\partial}{\partial y} (Ax - b) \end{pmatrix} = \begin{pmatrix} A^T \\ 0 \end{pmatrix}$$

$$r(x,y) \approx r(\bar{x},\bar{y}) + \left.\frac{\partial r}{\partial x}\right|_{(\bar{x},\bar{y})} \Delta x + \left.\frac{\partial r}{\partial y}\right|_{(\bar{x},\bar{y})} \Delta y$$

$$= \begin{pmatrix} \nabla_x f(\bar{x}) + A^T \bar{y} \\ A\bar{x} - b \end{pmatrix} + \begin{pmatrix} \nabla_{xx} f(\bar{x}) \\ A \end{pmatrix} \Delta x + \begin{pmatrix} A^T \\ 0 \end{pmatrix} \Delta y = 0$$

$$\underbrace{\quad\quad\quad}_{\substack{r(\bar{x},\bar{y}) \\ \text{(constant term)}}} \quad \underbrace{\quad\quad}_{\partial r/\partial x} \quad \underbrace{\quad\quad}_{\partial r/\partial y}$$

$$\rightarrow \quad \underbrace{\begin{pmatrix} \nabla_{xx} f(\bar{x}) & A^T \\ A & 0 \end{pmatrix}}_{\text{KKT matrix}} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} = - \underbrace{\begin{pmatrix} \nabla_x f(\bar{x}) + A^T \bar{y} \\ A\bar{x} - b \end{pmatrix}}_{\text{residual vector}}$$

Equality-constrained Newton method:

start with $(x^{(0)}, y^{(0)})$

for $k = 1 \rightarrow N_{max}$

$K = \text{ConstructKKT}(x^{(k-1)}, y^{(k-1)})$

$r = \text{ComputeResidual}(x^{(k-1)}, y^{(k-1)})$

$(\Delta x, \Delta y) = \text{SolveKKT}(K, r)$

$\alpha = \text{LineSearch}(\Delta x, \Delta y, \ldots)$

$x^k = x^{k-1} + \alpha \Delta x$

$y^k = y^{k-1} + \alpha \Delta y$

if $\|$ residual vector $(x^k, y^k) \|_2 < \varepsilon$

break

Complexity

KKT matrix: $\begin{pmatrix} \nabla_{xx} f(\bar{x}) & A^T \\ A & 0 \end{pmatrix} \in \mathbb{R}^{(m+n) \times (m+n)}$

since $A \in \mathbb{R}^{m \times n}$ and $\nabla_{xx} f \in \mathbb{R}^{n \times n}$

→ computing matrix inverse is $O((m+n)^3)$

Alternate derivation (feasible start Newton method)

Given $\bar{x}$ such that $A\bar{x}=b$

solve for $\Delta x \in \mathcal{N}(A)$ such that

$$A(\bar{x}+\Delta x) = A\bar{x} + \cancel{A\Delta x} = b$$

$$\min_{x} f(x) \quad \text{s.t. } Ax=b \quad \rightarrow \quad \min_{\Delta x} f(\bar{x}) + \nabla_x f(\bar{x})^T \Delta x + \frac{1}{2}\Delta x^T \nabla_{xx} f(\bar{x}) \Delta x$$

$$\text{subj. to: } A(x+\Delta x)=b \Leftrightarrow A\Delta x = 0$$

Taylor series expansion of $f(x)$ w.r.t. $\bar{x}$

$$\rightarrow \mathcal{L}(\Delta x, y) = f(\bar{x}) + \nabla_x f(\bar{x})^T \Delta x + \frac{1}{2}\Delta x^T \nabla_{xx} f(\bar{x}) \Delta x + \underbrace{y^T A \Delta x}_{A\Delta x = 0} = 0$$

$$\rightarrow \nabla_{\Delta x} \mathcal{L} = \nabla_x f(\bar{x}) + \nabla_{xx} f(\bar{x}) \Delta x + A^T y = 0$$

$$\nabla_y \mathcal{L} = A \Delta x = 0$$

$$\rightarrow \begin{pmatrix} \nabla_{xx} f(\bar{x}) & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ y \end{pmatrix} = - \begin{pmatrix} \nabla_x f(x) \\ 0 \end{pmatrix}$$

(feasible start Newton because $A\bar{x} = b$)

When is KKT matrix invertible?

$$\begin{pmatrix} \nabla_{xx} f(\bar{x}) & A^T \\ A & 0 \end{pmatrix}$$

1. A has full (row) rank
2. $\nabla_{xx} f(\bar{x})$ is positive definite $\forall x$

↳ all eigenvalues of Hessian must be positive