Math 164

Prof. W. Diepeveen | Fall 2024

Contents

1 Review

Calculus:

1. Jacobian of $f : \mathbb{R}^n \to \mathbb{R}^m$:

$$\boxed{\text{Jac}(f) = \left(\frac{\partial f_i}{\partial x_j}\right)_{ij}}$$

2. Hessian of $f : \mathbb{R}^n \to \mathbb{R}$:

$$\boxed{\nabla^2 f = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}\right)_{ij}}$$

3. ***Taylor expansion***: $f(y) = f(x) + \nabla f(x)^T (y - x) + \frac{1}{2}(y - x)^T \nabla^2 f(x)(y - x) + O(||y - x||^3)$
4. ***Directional derivative***: $\frac{v^T \nabla f(x)}{||v||}$

Quadratic forms: $\underline{f(x) = \frac{1}{2} x^T Q x - b^T x + c} \implies \underline{x^* = Q^{-1} b}$

- *Sylvester's criterion*: Q PD iff minors $d^k > 0 \; \forall \; k$; ND iff $(-1^k) d^k > 0 \; \forall \; k$

Symmetric matrices Q: can rewrite as $Q = V^T \Lambda V$

Convex sets: S convex $\Leftrightarrow \forall \; x, y \in S \; \& \; \alpha \in [0, 1]$, $\alpha x + (1 - \alpha) y \in S$

2 Basics of Optimization

Want to solve problems of the form:

$$\boxed{\min\ f(x) \text{ s.t. } f : \mathbb{R}^n \to \mathbb{R},\ \Omega \subseteq \mathbb{R}^n}$$

Def: A ***local minimizer*** of f under Ω is a point $x^* \in \Omega$ if $\exists\ \epsilon > 0$ s.t.

$$f(x^*) \leq f(x)\ \forall\ x \in \Omega \cap B_\epsilon(x^*)$$

Conditions for Local Minima

1. **FONC**:

$$x^* \text{ is a local min } \implies d^T \nabla f(x^*) \geq 0\ \forall\ d \in \mathbb{R}^n \text{ feasible}$$
$$[\Omega = \mathbb{R}^n : \nabla f(x^*) = 0]$$

2. **SONC**:

$$x^* \text{ is a local min, } d^T \nabla f(x^*) = 0 \implies d^T D_x^2 f(x^*) d \geq 0\ \forall\ d \in \mathbb{R}^n \text{ feasible}$$
$$[\Omega = \mathbb{R}^n : D_x^2 f(x) \geq 0]$$

3. **SOSC**:

$$x^* \text{ is an interior point, } \nabla f(x^*) = 0,\ D^2 f(x^*) > 0 \implies x^* \text{ is a local min}$$

3 1D Line Search

Goal: Want to minimize functions $f : \mathbb{R} \to \mathbb{R}$.

Golden Section/Fibonacci Search [0^{th} Order]

1. Start with search region $[a_0, b_0]$
2. At each step: pick $a_1, b_1 \in \mathbb{R}$ s.t. $a_1 - a_0 = b_1 - b_0 = \rho(b_0 - a_0)$ (for some ρ)
3. If $f(a_1) > f(b_1)$, pick a_1 as new left endpoint (replacing a_0); otherwise, pick b_1 as new right endpoint (replacing b_0).
4. Repeat steps 2 & 3.

Want to pick ρ intelligently s.t. either $b_k = a_{k+1}$ or $a_k = b_{k+1}$:

1. ***Golden Section***:

$$\rho^* = \frac{3 - \sqrt{5}}{2} \approx 0.382$$

2. ***Fibonacci Method***: For Fibonacci numbers $F_1, F_2, \ldots, F_N \in \mathbb{N}$

$$\rho_1 = 1 - \frac{F_N}{F_{N+1}},\ \rho_2 = 1 - \frac{F_M}{F_{N-1}},\ \ldots,\ \rho_N = 1 - \frac{F_1}{F_2}$$

Bisection Search [1^{st} Order]

1. Start with search region $[a_0, b_0]$
2. At every iteration, evaluate $f'(\frac{b_n + a_n}{2})$
3. $f' > 0 \implies$ choose $a_{n+1} = a_n, b_{n+1} = \frac{b_n + a_n}{2}$; else, choose $a_{n+1} = \frac{b_n + a_n}{2}, b_{n+1} = b_n$

Secant Method [1^{st} Order]

Have update rule:

$$x^{k+1} := x^k - \frac{x^k - x^{k-1}}{f'(x^k) - f'(x^{k-1})} f'(x^k)$$

Newton's Method [2^{nd} Order]

Have update rule:

$$x^{k+1} := x^k - \frac{f'(x^k)}{f''(x^k)}$$

[$(*)$ Requires $f''(x^k) > 0$ within the search region]

Performance

Golden Section < Fibonacci < Bisection Search < Secant Method < Newton's Method

Optimizing vs Zero-Finding

Many optimization algorithms can also be used to find the roots/zeroes of a function:

Optimizing: Finding zeroes of $f'(x)$

$\Updownarrow$

Zero-Finding: Finding zeroes of $f(x)$

4 Gradient Methods

4.1 Gradient Search

Fixed-Step-Size Gradient Descent

For some fixed step size $\alpha > 0$, have search direction $d^k = \nabla f(x^{(k)})$:

$$x^{(k+1)} := x^{(k)} - \alpha \nabla f(x^{(k)})$$

Steepest Descent

For each k, have update rule:

$$x^{(k+1)} := x^{(k)} - \alpha_k \nabla f(x^{(k)})$$

where:

$$\alpha_k = \arg\min_{\alpha \geq 0} f\left(x - \alpha \nabla f(x^{(k)})\right)$$

Convergence:

- "Order-p convergence":

$$0 < \lim_{k \to \infty} \frac{\left|\left|x^{k+1} - x^*\right|\right|}{||x^k - x^*||^p} < \infty \quad [\text{order-}\infty\text{: } \frac{||\cdot||}{||\cdot||^p} = 0 \ \forall \ p]$$

- For a quadratic form

$$f = \frac{1}{2} x^T Q x - b^T x + c; \ x^* = Q^{-1} b, \ \alpha_{opt} = \frac{\nabla f(x^k)^T \nabla f(x^k)}{\nabla f(x^k)^T Q \nabla f(x^k)}$$

  - Fixed-step-size GD converges if:

$$0 < \alpha < \frac{2}{\lambda_{min}(Q)}$$

  - Steepest descent converges always

- Convergence rate is worst-case linear

4.2 Convergence Proofs

Fixed SS GD: Alg satisfies $V(x^{k+1}) = (1 - \gamma_k) V(x^k)$, where $\gamma_k = 1$ if $g^k = 0$, otherwise

$$\gamma_k = \alpha_k \frac{(g_k)^T Q g^k}{(g^k)^T Q^{-1} g^k} \left(2 \frac{(g^k)^T g^k}{(g^k)^T Q g^k} - \alpha\right)$$

Theorem. x^k, γ_k as above, $\gamma_k > 0$; then $x^k \to x^*$ for any x^0 iff $\sum_{k=0}^{\infty} \gamma_k = \infty$.

Use that $\lambda_{\min}(Q)\,||x||^2 \leq x^T Q x \leq \lambda_{\max}(Q)\,||x||^2$; $\lambda_{\min}(Q^{-1}) = 1/\lambda_{\max}(Q)$

5 Newton's Method + Variations

Newton's Method ($\mathbb{R}^n$)

$$x^{(k+1)} := x^{(k)} - (D^2 f(x^{(k)}))^{-1} \nabla f(x^{(k)})$$

Convergence:

- $d^{(k)}$ only guaranteed to be a descent direction if $D^2 f(x^{(k)} > 0$
- No guarantee that $f(x^{(k+1)}) < f(x^{(k)})$
- Criteria: If $f \in C^3$ and $x^* \in \mathbb{R}^n$ s.t. $\nabla f(x^*) = 0$ & $D^2 f(x^*)$ invertible, then Newton's method converges with order ≥ 2 in some neighborhood of x^*
  - For quadratic f, converges in a single step
- Can use line search initially to find a better starting point $x^{(0)}$

Convergence Proof: Derive inequalities & look at $||x^1 - x^*||$, take $||x^0 - x^*|| \leq \alpha/(c_1 c_2)$

1. $\nabla f(x) - \nabla f(x^0) - D^2 f(x^0)(x - x_0) = O(||x - x_0||^2) \leq c_1 \, ||x - x_0^2||$ [Via Taylor]
2. From regularity: for $x \in B_\epsilon(x^*)$, have that $||(D^2 f(x))^{-1}|| \leq c_2$ for some $c_2 \in \mathbb{R}$

5.1 Modifications of Newton's Method

Levenberg-Marquardt Algorithm

For more stability, can use update rule [with $\mu > -\lambda_{min}(D^2 f(x^{(k)}))$; new eigenvalues $\lambda_i + \mu$]:

$$x^{(k+1)} := x^{(k)} - (D^2 f(x^{(k)}) + \mu I_n)^{-1} \nabla f(x^{(k)})$$

Note:

$$||d^k|| \leq \frac{1}{\lambda_{\min}} ||\nabla f(x^k)|| \implies ||d^k|| \leq \frac{1}{\lambda_{\min} + \mu} ||\nabla f(x^k)||$$

Gauss-Newton Algorithm

When optimizing nonlinear least squares problems of the form:

$$f(x) = \sum_{i=0}^{m} (r_i(x))^2;\ r_i : \mathbb{R}^n \to \mathbb{R},\ r = \langle r_1,\ r_2,\ \ldots,\ r_m \rangle$$

For more efficiency, if we expect $r_i(x^*) \approx 0$, can use update rule (for $J = \nabla r$):

$$x^{(k+1)} := x^{(k)} - (J^T J)^{-1} J r(x^{(k)}))$$

6 Other Optimization Methods

6.1 Conjugate Direction Methods

Definition. For quadratic form $f(x) = \frac{1}{2}x^TQx - x^Tb, Q = Q^T > 0$: a set of directions $d^1, d^2, \ldots, d^m$ are called ***Q-conjugate*** if $(d^i)^TQd^j = 0 \; \forall \; i \neq j$. [Note: Q-conjugate $\implies$ linearly independent]

The Conjugate Direction Algorithm

Given Q-conjugate directions $d^1, \ldots, d^k$:

$$\underline{x^{k+1} := x^k + \alpha_k d^k} \text{ where } g^k = Qx^k - b, \; \alpha_k = \frac{(g^k)^T d^k}{(g^k)^T Q g^k} \text{ [stop if } g^{k+1} = 0]$$

- For any $x^0 \in \mathbb{R}^n$, given n Q-conjugate directions, the alg. converges to $x^* = Q^{-1}b$ in at most n steps.
- ***Conjugate Gradient Algorithm***: Find Q-conjugate directions by solving:

$$\underline{\beta_k = \frac{(g^{k+1)^TQd^k}}{(d^k)^TQd^k}} \implies \underline{d^{k+1} := -gk+1] + \beta_k d^k} \quad [d^0 =]$$

6.2 Quasi-Newton Methods

Idea: $x^{k+1} := x^k - \alpha_k H_k \nabla f(x^k)$, $\alpha_k = \arg\min_{\alpha \geq 0} f(x^k + \alpha H_k \nabla f(x^k))$ for some approximation $H_k \approx (D^2 f(x^k))^{-1}$

- Impose constraint $\forall \; k$: $H_{k+1}\nabla g^i = \nabla x^i$ for $i = 1, \ldots, k$ [note: these are conj dir methods]
  After n steps: obtain n linear equations & solve [if ∇G^n nonsingular]:

$$H_n \nabla G^n = \nabla X^n \to H_n = \nabla X^n (\nabla G^n)^{-1}$$

$\implies$ Solution unique, Q^{-1} a soln. $\implies$ $H_n = Q^{-1}$; yields Q^{-1} after n iterations, converges on $(n+1)^{th}$

Rank-One Correction: Keep adding "degrees of freedom" $(rank[z^k(z^k)^T] = 1)$

$$H_{k+1} := H_k + a_k z^k (z^k)^T \; [z^k \in \mathbb{R}^n], \; \text{where} a_k z^k (z^k)^T = \frac{(\nabla x^k - H_k g^k)(\nabla x^k - H_k \nabla g^k)^T}{(\nabla g^k)^T(\nabla x^k - H_k \nabla g^k)} \text{ [from constraint]}$$

DFP: Given $x^0 \in \mathbb{R}^n$, H_0 any symmetric PD matrix (e.g. $I_{n \times n}$:

$$H_{k+1} := H_k + \frac{\nabla x^k (\nabla x^k)^T}{(\nabla x^k)^T \nabla g^k} - \frac{H_k \nabla g^k (H_k \nabla g^k)^T}{(\nabla g^k)^T H_k \nabla g^k}$$

Lemma (Sherman-Morris)**.** Let $A \in \mathbb{R}^{n \times n}$ nonsingular & $u, v \in \mathbb{R}^n$ s.t. $1 + V^t A u \neq 0$; then:

$$(A + u^T v)^{-1} = A^{-1} - \frac{(A^{-1}u)(v^T A^{-1})}{1 + v^T A u}$$

BFGS: Approximate $D^2 f(x)$ and take inverse of approximation

$$B_{k+1} = B_k - \frac{\nabla x^k (\nabla x^k)^T}{(\nabla x^k)^T \nabla g^k} - \frac{B_k \nabla g^k (B_k \nabla g^k)^T}{(\nabla g^k)^T B_k \nabla g^k} \quad \implies \quad H_{k+1} = (B_{k+1})^{-1}$$

7 Solving Linear Systems

Want to solve (for $A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m$):

$$\boxed{x \in \mathbb{R}^n \text{ s.t. } Ax = b}$$

Three cases:

1. $\underline{m \geq n, \text{rank}\,(A) = m}$: May not be a solution; pick:

$$x^* = \min_{x\in\mathbb{R}^n} ||Ax - b||^2 \implies \boxed{x^* = (A^TA)^{-1}A^Tb}$$

2. $\underline{m \leq n, \text{rank}\,(A) = n}$: Infinitely many solutions, pick:

$$x^* = \min_{x\in\mathbb{R}^n} ||x|| \text{ s.t. } Ax = b \implies \boxed{x^* = A^T(AA^T)^{-1}b}$$

3. $\underline{\text{rank}\,(A) = r \leq \min\{m, n\}}$: Solve via pseudoinverse $A^\dagger$

Kaczmarz algorithm: Solves $m \leq n$ case $[x^k \to x^*]$ w/ purely rank 1 updates

Lemma (Full-rank factorization)**.** Let $A \in \mathbb{R}^{m\times n}$ & $\text{rank}\,(A) = r$; then $\exists$ matrices $B \in \mathbb{R}^{m\times r}, C \in \mathbb{R}^{r\times m}$ s.t. $A = BC$ $[\text{rank}\,(b) = \text{rank}\,(C) = r]$

Definition. Given $A = \mathbb{R}^{m\times n}$; a matrix $A^\dagger \in \mathbb{R}^{n\times m}$ is called a [the] ***pseudo-inverse*** of A if:

$$\boxed{AA^\dagger A = A}$$

and $\exists$ matrices $U \in \mathbb{R}^{n\times n}, V \in \mathbb{R}^{m\times m}$ satisfying $A^\dagger = UA^T = A^TV$. [Note: always exists + unique]

- If A invertible, then $A^\dagger = A^{-1}$ with $U = (A^TA)^{-1}$, $V = (AA^T)^{-1}$
- In each of 2 cases above, solution is given by $A^\dagger$

Recursive Least Squares: Assume have solution x^0 for $A_0x = b_0$, but obtain new measurements A_1, b_1 & want to update; $G_1 = \begin{smallmatrix} A_0\,T & A_0 \\ A_1 & A_1 \end{smallmatrix} = G_0 + A_1^TA_1 \implies x^1 = x^0 + G_1^{-1}A_1^T(G^1 - A_1x_0)$ Sherman-Morrison-Woodbury: A nonsingular, U, V matrices s.t. $1 + VA^{-1}U \neq 0$; then:

$$(A + UV)^{-1} = A^{-1} - (A^{-1}U)(1 + VA^{-1}U)(VA^{-1})$$

$$\implies P_{k+1} = P_kA_{k+1}^T(1 + A_{k+1}P_kA_{k+1}^T)A_{k+1}P_k \implies x_{k+1} = x^k + P_{k+1}A_{k+1}^T(b^{k+1} - A_{k+1}x^k)$$

$$\implies P_{k+1} = P_k - \frac{P_ka_{k+1}a_{k+1}^TP_k^T}{1 + a_{k+1}^TP_ka_{k+1}} \implies x_{k+1} = x_k + P_{k+1}a_{k+1}(b_{k+1} - a_{k+1}^Tx_k) \text{ [rank 1 } A_1]$$

8 Linear Programming

Linear programming (standard form): want to find [for $c \in \mathbb{R}^n, A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$; $\text{rank}\,(A) = m$]:

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

Correspond to optimization problems over convex polytopes $\Omega = \{x \in \mathbb{R}^n : Ax = b\}$

Definition. Let $B \in \mathbb{R}^{m \times m}$ be constructed from any m linearly independent columns of A; then the vector $[B^{-1}b \;\; 0]^T \in \mathbb{R}^n$ [solving $Ax = b$] is called a ***basic solution*** w.r.t. basis B.

- If $x_B = B^{-1}b$ has entries with value 0, call it a ***degenerate basic solution***
- If $x_B \geq 0$, call it a ***basic feasible solution***
- Basic solutions are equivalent to "extreme points" [vertices] of convex polytope Ω

Fundamental Theorem of LP: for any LP:

1. If $\exists$ a feasible solution, then $\exists$ a basic feasible solution
2. If $\exists$ an optimal feasible solution, then $\exists$ an optimal basic feasible solution

Simplex Algorithm: Algorithm for solving standard form LPs via moving between adjacent corners of Ω

- Given an LP and initial basic feasible solution x: the simplex algorithm either returns an optimal solution x^* (if one exists), or finds that the LP is unbounded
- ***Two-phase simplex method***: To find initial point x, can solve associated "artificial problem":

$$\text{minimize} \;\; y_1 + \ldots + y_m \text{ subject to } \begin{bmatrix} A & I_m \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = b, \quad x, y \geq 0$$

  $\implies$ original LP has a basic feasible soln iff artificial problem has an optimal feasible soln. with $y = 0$

Simplex:

1. Swapping bases: $a_q = y_{10}a_1 + \ldots + y_{m0}a_m \implies \exists\, y_{1q}, \ldots, y_{mq}$ s.t.

$$\forall\, \epsilon > 0 : \; (y_{10} - \epsilon y_{1q})a_1 + \ldots + (y_{m0} - \epsilon y_{mq})a_m + \epsilon a_q = b; \quad \text{pick } \epsilon = \min_i \left\{ \frac{y_{i0}}{y_{iq}} : y_{iq} > 0 \right\}$$

8.1 Duality

Duality (Symmetric form):

$$\left[\textbf{\textit{Primal}}\right] \quad \left.\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \geq b \\ & x \geq 0 \end{array}\right\} \iff \left\{\begin{array}{lll} \text{maximize} & \lambda^T b & \\ \text{subject to} & \lambda^T A \leq c^T & \left[\textbf{\textit{Dual}}\right] \\ & \lambda \geq 0 & \end{array}\right.$$

Duality (Asymmetric form):

$$\left[\textbf{\textit{Primal}}\right] \quad \left.\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \geq 0 \end{array}\right\} \iff \left\{\begin{array}{lll} \text{maximize} & \lambda^T b & \\ \text{subject to} & \lambda^T A \leq c^T & \left[\textbf{\textit{Dual}}\right] \end{array}\right.$$

Symmetric form: dual of the dual is the primal [not true for asymmetric]

Theorem (**Weak Duality**)**.** Let $x \in \mathbb{R}^n, \lambda \in \mathbb{R}^m$ be feasible solutions to the primal and dual LP problems, respectively (either form). Then:

$$c^T x \geq \lambda^T b$$

Theorem (**Duality Theorem/Strong Duality**)**.** If the primal LPproblem (in either form) has an optimal solution, then so does the dual problem, and the optimal values for the objective functions are equal for both problems.

Theorem (**Complementary Slackness**)**.** A pair of feasible solutions $x \in \mathbb{R}^n, \lambda \in \mathbb{R}^m$ is optimal iff:

1. $(c^T - \lambda^T A_x = 0$
2. $\lambda^T (Ax - b) = 0$

9 Nonlinear Programming

For $x \in \mathbb{R}^n, f : \mathbb{R}^n \to \mathbb{R}, h : \mathbb{R}^n \to \mathbb{R}^m, g : \mathbb{R}^n \to \mathbb{R}^p$:

$$\begin{array}{rl} \text{minimize} & f(x) \\ \text{subject to} & h(x) = 0 \\ & g(x) \leq 0 \end{array}$$

Definition.

1. A constraint $g_i(x^*)$ is said to be ***active*** at $x^* \in \mathbb{R}^n$ if $g_i(x^*) = 0$. [All h_i are ***active***.]
2. Define the ***tangent space*** [generated by the active constraints] at x^* by:

$$T(x^*) = \left\{ y \in \mathbb{R}^n : y^T \nabla h(x^*) = 0,\ y^T \nabla g_i(x^*) = 0 \ \forall\ g_i \text{ active} \right\}$$

3. Call x^* (feasible) a ***regular point*** if:

$$\underline{\{\nabla h_i(x^*) : i = 1, \ldots, p\} \cup \{\nabla g_j(x^*) : g_j \text{ active}\} \text{ is linearly independent}}$$

Definition. Define the ***Lagrangian*** as the function $\ell : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ given by:

$$\ell(x, \lambda, \mu) = f(x) + h(x)^T \lambda + g(x)^T \mu$$

$$\Downarrow$$

1^{st}-Order Necessary Conditions (KKT)

Let x^* be a local min of $f : \mathbb{R}^n \to \mathbb{R}$ s.t. $h(x^*) = 0, g(x^*) \leq 0$ for $h : \mathbb{R}^n \to \mathbb{R}^m, g : \mathbb{R}^n \to \mathbb{R}^p$. Assume x^* is a regular point; then $\exists\ \lambda^*$ [***Lagrange multi.***], μ^* [***KKT multi.***] such that:

1. $\mu^* \geq 0$
2. $\nabla f(x^*) + \nabla h(x^*)^T \lambda^* + \nabla g(x^*)^T \mu^* = 0 \quad [\text{i.e.} \nabla_x \ell(x, \lambda, \mu) = 0]$
3. $(\mu^*)^T (g(x^*)) = 0$

2^{nd}-Order Conditions (SONC & SOSC)

For x^*, λ^*, μ^* points (x^* a regular point) satisfying the FONC and $f, g, h \in \mathcal{C}^2(\mathbb{R}^n)$:

1. **SONC**: $y^T D\ell(x^*, \lambda^*, \mu^*) y \geq 0 \ \forall\ y \in T(x^*) \quad [D\ell(x^*, \lambda^*, \mu^*) \text{ is } \underline{\text{P.S.D.}} \text{ on } T(x^*)]$
2. **SOSC**: $y^T D\ell(x^*, \lambda^*, \mu^*) y > 0 \ \forall\ y \in T(x^*) \quad [D\ell(x^*, \lambda^*, \mu^*) \text{ is } \underline{\text{P.D.}} \text{ on } T(x^*)]$

- Notice: $\mu_i^* = 0 \ \forall\ i$ for which $g_i(x^*)$ is active