Allen Zhu

Linear algebra

Norms satisfy

- $\|x\| \geq 0$ with equalify iff $x = 0$
- $\|ax\| = |a| \, \|x\|$
- Triangle: $\|x + y\| \leq \|x\| + \|y\|$

Dual norm $\|x\|_* = \max_{\|z\| \leq 1} z^T x$

Trace. $\mathrm{tr}(ABC) = \mathrm{tr}(CAB)$

Complex conjugate. For real U, $U^* = U^T$

Orthonormal columns: $U^T U = I$

Orthogonal (unitary) matrix: $U^T U = UU^T = I$ or $U^{-1} = U^T$
Columns and rows are orthonormal

Diagonalizable matrix $A = PDP^{-1}$

Symmetric (normal) matrix: $A = A^T$ $(AA^* = A^* A)$
Spectral theorem. $A = U \Lambda U^T$, U orthonormal
Singular values are $|\Lambda|$ since $A^2 = U \Sigma^2 U^T$
SVD $A = U \Sigma V^T$ where $v_i = \mathrm{sign}(\lambda_i) u_i$: $\sum u_i \lambda_i u_i^T = \sum u_i |\lambda_i| \mathrm{sign}(\lambda_i) u_i^T$

Singular values are square roots of nonnegative eigenvalues of $A^T A$

SVD: $X = U \Sigma V^T$. If X has $m > n$, U and V have orthogonal columns. U is $m \times n$; Σ and V are $n \times n$.

Operator norm. $\|A\|_{op} = \max_{\|v\|=1} \|Av\|$. $\|Av\| \leq \|A\|_{op} \|v\|$, $\|AB\| \leq \|A\| \|B\|$

Operator norm for ℓ_2 vector space is the max singular value
For any $x = \sum c_i v_i$, $\frac{|Ax|^2}{|x|^2} = \frac{\sum c_i^2 \lambda_i^2}{\sum c_i^2} \leq \max \lambda_i^2$.

Frobenius norm $\sqrt{\mathrm{tr}(A^T A)}$ is norm based on element-wise dot product

$Av = \lambda v \Rightarrow (A + cI)v = (\lambda + c)v$

$u^T A v = \sum_{i,j} u_i v_j A_{ij}$ and $ABC_{ij} = \sum_{m,n} A_{im} B_{mn} C_{nj}$

Topology

Compact = closed and bounded

Continuous functions preserve compact sets

Neighborhood of p of radius r $N_r(p) = \{q : |p - q| < r\}$

p is an *interior point* $\Leftrightarrow \exists\delta$ s.t. $N_\delta(p) \subset S$

p is a *boundary point* $\Leftrightarrow \forall r$, $N_r(p)$ contains $p_1 \in S$ and $p_2 \notin S$

p is a *limit point* of $S \Leftrightarrow$ every neighborhood of p contains $q \neq p$ s.t. $q \in S$ $\Leftrightarrow p$ is the limit of a sequence of points in S

S open $\Leftrightarrow$ every point in S is an interior point

S closed $\Leftrightarrow$ S contains every limit point of S $\Leftrightarrow$ S contains its boundary $\Leftrightarrow S^c$ is open

Union of open sets is open, intersection of finite open sets is open

$\emptyset$ and $\mathbb{R}$ are both open and closed; $(0, 1]$ is neither open or closed

A *topology* or *topological space* is defined by a set and a choice of open subsets satisfying the axioms. Every set is open in the discrete topology. No set except $\emptyset$ and the space itself is open in the indiscrete topology.

Analysis

f is *continuous* at a iff $\forall\varepsilon > 0,\ \exists\delta > 0$ s.t. $|x - a| < \delta \Rightarrow |f(x) - f(a)| < \varepsilon$

Cauchy-Schwartz inequality. $\|u \cdot v\| \leq \|u\|\|v\|$

Mean value theorem. $\exists z \in (x, y)$ s.t. $f(y) - f(x) = \nabla f(z)^T (y - x)$

Given the 1D version, $f'(z) = \frac{f(y)-f(x)}{y-x}$, we can define $g(t) = f((1-t)x + ty)$. Then $\exists c$ s.t. $g(1) - g(0) = g'(c)$.

Mean value form of Taylor's theorem.

$$\exists z \in (x, y) \text{ s.t. } f(y) = f(x) + f'(x)(y - x) + \frac{1}{2} f''(z)(y - x)^2$$

Hessian-vector product. $\nabla^2(x)\mathbf{v} = \lim_{h\to 0} \frac{\nabla f(x + h\mathbf{v}) - \nabla f(x)}{h}$

Allen Zhu

Convexity

Note: all statements involving t hold for all $t \in (0,1)$

Set S is *convex* iff $tx + (1-t)y \in S \ \forall x, y \in S$

0° f is **convex** iff its domain is convex and

$$f(tx + (1-t)y) \le tf(x) + (1-t)f(y) \ \forall x, y$$

f is *strictly convex* iff the inequality is strict

f is *strongly convex* for param $m > 0$ iff $f - \frac{m}{2}\|x\|_2^2$ is convex

Indicator function of a convex set S is convex: $f(x) = 0$ if $x \in S$ else ∞

Quadratic $\frac{1}{2}x^T Qx + b^T x + C$ is convex iff $Q \succeq 0$

f convex iff its *epigraph* $\{(x,t) | f(x) \le t\}$ is convex

If f is convex or quasiconvex, all its *sublevel sets* $\{x | f(x) \le t\}$ are convex

1° **Lower Linear Bound**. f is convex iff its domain is convex and

$$f(y) \ge f(x) + \nabla f(x)^T (y - x) \ \forall x, y$$

Cor. x is a global minimum iff $\nabla f(x)^T (y-x) \ge 0 \ \forall y$. (Boyd 139)

Cor. x is a global minimum if $\nabla f(x) = 0$ (or $\exists 0 \in \partial f(x^*)$)

Convex $\Rightarrow 1^\circ$. *Pf.* $\forall x, y$, we have

$$\begin{aligned} tf(y) + (1-t)f(x) \ge f(ty + (1-t)x) &= f(x + t(y-x)) \\ f(y) \ge f(x) &+ \frac{f(x + t(y-x)) - f(x)}{t} \\ &\ge f(x) + \nabla(x)^T (y-x) \text{ after } t \to 0 \end{aligned}$$

$1^\circ \Rightarrow$ convex. *Pf.* Let $z = tx + (1-t)y$.

$$\begin{aligned} f(x) &\ge f(z) + \nabla f(z)^T (x - z) \\ f(y) &\ge f(z) + \nabla f(z)^T (y - z) \\ tf(x) + (1-t)f(y) &\ge f(z) + \nabla f(z)^T (tx + (1-t) - z) \\ tf(x) + (1-t)f(y) &\ge f(z) \end{aligned}$$

2°. f is convex iff its domain is convex and

$$\nabla^2 f(x) \succeq 0 \ \forall x$$

$\nabla^2 f(x) \succ 0 \Rightarrow f$ is strictly convex but the converse is not true (x^4 at $x = 0$)

Convex $\Rightarrow \nabla^2 \succeq 0$ (1D). *Pf.* For $x < y$, we have $f(y) \geq f(x) + f'(x)(y - x)$ and $f(x) \geq f(y) + f'(y)(x - y)$, so $f'(x)(y - x) \leq f(y) - f(x) \leq f'(y)(y - x)$. Dividing by $(y - x)^2$, $\frac{f'(y) - f'(x)}{y - x} \geq 0$. So as $y \to x$, $f''(x) \geq 0$.

$\nabla^2 \succeq 0 \Rightarrow$ convex (1D). *Pf.* By the mean value form of Taylor's theorem, $\exists z \in [x, y]$ s.t. $f(y) = f(x) + f'(x)(y - x) + \frac{1}{2} f''(z)(y - x)^2$, so $f(y) \geq f(x) + f'(x)(y - x)$.

Generalization to $\mathbb{R}^n$. *Pf.* f is convex $\Leftrightarrow f$ is convex along all lines $\Leftrightarrow g(r) = f(x + rv)$ is convex for all $x, v \Leftrightarrow g''(r) = v^T \nabla^2 f(x + rv) v \geq 0 \Leftrightarrow \nabla^2 f(x) \succeq 0$.

Monotonicity condition. f is convex iff its domain is convex and

$$(\nabla f(x) - \nabla f(y))^T (x - y) \geq 0$$

Convex $\Rightarrow$ monotonic. *Pf.* Apply the first-order characterization to (x, y) and (y, x) and combine the inequalities:

$$\begin{aligned} f(y) &\geq f(x) + \nabla f(x)^T (y - x) \\ f(x) &\geq f(y) + \nabla f(y)^T (x - y) \\ \nabla f(x)^T (x - y) &\geq \nabla f(y)^T (x - y) \end{aligned}$$

Monotonic $\Rightarrow$ convex. *Pf.* TODO

Partial optimization: $\min_{x_i} f$ is convex for any subset of variables x_i

Smoothness

f is L-smooth iff:

- 0° **Upper quadratic bound**. $f(y) \leq f(x) + \nabla f(x)^T (y - x) + \frac{L}{2} \|y - x\|_2^2$
- 1° $\langle \nabla f(x) - \nabla f(y), x - y \rangle \leq L \|x - y\|_2^2$
- 2° $\nabla^2 f(x) \preceq LI$

Allen Zhu

Gradient descent

GD Lemma. Function value decreases each iteration. Can still take forever since $\nabla f \to 0$ as $x \to x^*$.

Plug GD def into smoothness and assume $\eta \leq \frac{1}{L}$:

$$\begin{aligned}
f(x_{t+1}) &\leq f(x_t) + \nabla f(x_t)^T(x_{t+1} - x_t) + \frac{L}{2}\|x_{t+1} - x_t\|_2^2 \\
f(x_{t+1}) &\leq f(x_t) + \nabla f(x_t)^T(-\eta \nabla f(x_t)) + \frac{L}{2}\| - \eta \nabla f(x_t)\|_2^2 \\
f(x_{t+1}) &\leq f(x_t) - \eta\|\nabla f(x_t)\|^2 + \frac{L\eta^2}{2}\|\nabla f(x_t)\|_2^2 \\
f(x_{t+1}) &\leq f(x_t) - \frac{\eta}{2}\|\nabla f(x_t)\|_2^2
\end{aligned}$$

MD Lemma. Use convex property and the identity $\langle x_t - x_{t+1}, y - x_t\rangle = -\frac{1}{2}\left(\|y - x_t\|^2 - \|y - x_{t+1}\|^2 + \|x_{t+1} - x_t\|^2\right)$

$$\begin{aligned}
f(y) &\geq f(x_t) + \langle \nabla f(x_t), y - x_t\rangle \\
&\geq f(x_t) + \frac{1}{\eta}\langle x_t - x_{t+1}, y - x_t\rangle \\
&\geq f(x_t) - \frac{1}{2\eta}\left(\|y - x_t\|^2 - \|y - x_{t+1}\|^2 + \|x_{t+1} - x_t\|^2\right) \\
f(x_t) &\leq f(y) + \frac{1}{2\eta}\left(\|y - x_t\|^2 - \|y - x_{t+1}\|^2 + \|x_{t+1} - x_t\|^2\right) \\
\sum f(x_t) &\leq Tf(x^*) + \frac{1}{2\eta}\left(\|y - x_t\|^2 - \|y - x_{t+1}\|^2 + \|x_{t+1} - x_t\|^2\right) \\
\frac{1}{T}\sum f(x_t) &\leq f(x^*) + \frac{1}{2\eta T}\|x^* - x_0\|^2 + \frac{\eta}{2T}\sum \|\nabla f(x_t)\|^2
\end{aligned}$$

Conditioning

Condition number $\kappa = \frac{\lambda_{\max}}{\lambda_{\min}}$.

$f(x) = f(x^*) + \cancel{\nabla f(x^*)}(x - x^*) + \frac{1}{2}(x - x^*)^T \nabla^2 f(z)(x - x^*)$, so $\nabla f(x) = H(x - x^*)$ for $H = \nabla^2 f(z)$. Let H have eigenvalues λ_i.

$$\begin{aligned} x_{t+1} &= x_t - \eta \nabla f(x_t) \\ x_{t+1} &= x_t - \eta H(x - x^*) \\ (x_{t+1} - x^*) &= (I - \eta H)(x_t - x^*) \\ \|\tilde{x}_{t+1}\|_2 &\leq \|I - \eta H\|_2 \|\tilde{x}_t\|_2 \end{aligned}$$

$1 - \eta H$ has eigenvalues $1 - \eta\lambda_i$.

For $\eta = \frac{c}{\lambda_{\max}}$, $0 < c < 2$, $\max|1 - \eta\lambda_i| = 1 - c\frac{\lambda_{\min}}{\lambda_{\max}} = 1 - \frac{c}{\kappa} < 1$.

For $\eta = \frac{2}{\lambda_{\min}+\lambda_{\max}}$, $\max|1 - \eta\lambda_i| = 1 - \frac{2\lambda_{\min}}{\lambda_{\min}+\lambda_{\max}} = \frac{\kappa-1}{\kappa+1}$.

Integrating, $\|\tilde{x}_{t+1}\| \leq \left(\frac{\kappa-1}{\kappa+1}\right)^t \|\tilde{x}_0\|$.