Matthew Tran
EE127
MT1

- Optimization
  - standard form
    $p^* = \min_x f_0(x)$
    subject to $f_i(x) \le 0$, $i=1,\dots,m$
- affine - linear + const, $a^Tx + b \to y$
  - feasible set - satisfy constraints
  - suboptimal set - $\epsilon$ within of $p^*$
  - level sets
  - sub-level set
  - equality reduction
    $h(x) = 0$
    $\Downarrow$
    $h(x) \le 0$, $h(x) \ge 0$
  - set constraint - $x \in \mathcal{X}$ (subset $\mathbb{R}^n$)
  - $\max \to \min$, $g_0(x) \to -g_0(x)$, $-p^*$
  - infeasible - empty feasible set
    - feasibility problem $f_0(x)$ constant
  - optimal set - argmin $f_0(x)$
    - empty if infeasible
    - empty if optimal is at limit
    - end on or < boundary
  - local min - optimal within R
  - focus on convex tractable problems
  - log transform $x_1 x_2 \to \log x_1 + \log x_2$
  - convex - local min == global min
- Vectors
  - represent point, direction, linear function
  - scalar product - "angle"
  - column format $\begin{pmatrix} x_1 \\ x_n \end{pmatrix} \in \mathbb{R}^n$
  - vector space - lin. combos
  - subspace - closed under +, scalar $\times$
  - span - all linear combo of S vectors
  - combine subspace $X \oplus Y \to x + y$ form
  - linearly independent
    $\sum \lambda_i x^{(i)} = 0 \Rightarrow \lambda = 0$ "dimension"
    basis of $S \to$ span$(B) = S$
  - affine set - $A = \{x \in X : x_0 + v^{(i)}, v \in V\}$
    - subspace $V$ is affine w/ origin [illegible]
    - line $L = \{x : x = x_0 + tv, v \in \{x_0 : x \in B\}\}$
  - Euclidean length - $\ell_2$ norm
  - norm
    $\|x\| \ge 0$ ; $\|x\| = 0$ iff $x = 0$
    $\|x+y\| \le \|x\| + \|y\|$ (tri. ineq.)
    $\|\alpha x\| = |\alpha| \|x\|$
  - $\ell_p$ norm $\|x\|_p = \left(\sum |x_k|^p\right)^{1/p}$
    - $\|x\|_2 = \sqrt{x^Tx}$
    - $\|x\|_1 = \sum |x_k|$  $\ell_0$ is cardinality, num nonzero
    - $\|x\|_\infty = \max |x_k|$
  - inner product $\langle x, y \rangle$
    $\langle x, x \rangle \ge 0$
    $\langle x, x \rangle = 0$ iff $x = 0$
    $\langle x+z, y \rangle = \langle x, y \rangle + \langle z, y \rangle$
    $\langle \alpha x, y \rangle = \alpha \langle x, y \rangle$
    $\langle x, y \rangle = \langle y, x \rangle$
    - inner product space w/ norm
    - standard inner product $\langle x, y \rangle = x^Ty$
    - induces a norm $\|x\| = \sqrt{\langle x, x \rangle}$
  - angle
    $\cos\theta = \frac{x^Ty}{\|x\|_2 \|y\|_2}$
    $x^Ty = 0 \to 90°$, orthogonal $x \perp y$
  - Cauchy-Schwartz $|x^Ty| \le \|x\|_2 \|y\|_2$
  - Hölder inequality
    for $p, q \ge 1$, $\frac{1}{p} + \frac{1}{q} = 1$,
    $|x^Ty| \le \sum_{k=1}^n |x_k y_k| \le \|x\|_p \|y\|_q$
  - mutually orthogonal $\to$ linearly independent
  - orthogonal decomposition
    $X = S \oplus S^\perp$
- Projection
  - min distance to set to point
  - proj theorem - unique proj $x^*$
    $(x - x^*) \perp S$
  - point to line $p^* = x_0 + v^T(p - x_0)v$
  - hyperplane $a^Tx = b$
    $p^* = p - \frac{a^Tp - b}{\|a\|_2^2} a$
  - orthonormal basis $x^* = \sum \langle u^{(i)}, x \rangle$
- Function - Map
  $f: \mathbb{R}^n \to \mathbb{R}$   $f: \mathbb{R}^n \to \mathbb{R}^m$
  graph $f = \{(x, f(x)) \in \mathbb{R}^{n+1} ; x \in \mathbb{R}^n\}$
  epi $f = \{(x, t) \in \mathbb{R}^{n+1} : x \in \mathbb{R}^n, t \ge f(x)\}$
  "everything above"
- affine $f(x) = a^Tx + b$
- linear $f(x) = a^Tx$
- halfspace $H_- : a^Tx \le b$
  $H_{++} : a^Tx > b$
- gradient
  $\nabla f(x) = \left[\frac{\partial f}{\partial x_1} \cdots \frac{\partial f}{\partial x_n}\right]^T$
  $f(x) \approx f(x_0) + \nabla f(x_0)^T(x - x_0)$
  1st order Taylor approx
  $\perp$ to contour level set, max ascent
- Matrix
  - linear map between spaces
  - $A \in \mathbb{R}^{m \times n}$, $m$ rows, $n$ columns
  $(A)_{ij}$ entries of element
  - product $[AB]_{ij} = \sum_{k=1}^n A_{ik} B_{kj}$
    row $\cdot$ column
  - $Ab = \sum_{k=1}^n a_k b_k$
  - $c^TA = \sum_{k=1}^m c_k a_k^T$ rows
  - $AB = \sum_{i=1}^n a_i b_i^T$ dyad (rank 1)
  - $(AB)^T = B^TA^T$
  - range $R(A) = \{Ax : x \in \mathbb{R}^n\}$
    column space
    $\dim$ rank$(A)$: rank$(A^T) \le \min(m, n)$
  - nullspace $N(A) = \{x \in \mathbb{R}^n : Ax = 0\}$
    $\mathbb{R}^n = N(A) \oplus R(A^T)$
    decomposes to sum of orthogonal vectors
  - Fundamental Theorem of Linear Algebra
    $R(A)^\perp = N(A^T)$
  - Determinant
    $\det(A) = \sum_{i=1}^n (-1)^{i+j} a_{ij} \det A_{(ij)}$
    A singular iff $\det A = 0$ or $N(A) \neq 0$
    $\det A = \det A^T$, $\det AB = \det BA = \det A \det B$
    $\det \alpha A = \alpha^n \det A$
  - Inverse
    - $AA^{-1} = A^{-1}A = I$
    - $AB^{-1} = B^{-1}A^{-1}$ ; $(A^T)^{-1} = (A^{-1})^T$ ; $\det A^{-1} = \frac{1}{\det A}$
    - pseudoinverse $AA^\dagger A = A$
  - Eigen
    - $Av = \lambda v$, $v \neq 0$  $(\lambda I - A)v = 0$
    - $p(\lambda) = \det(\lambda I - A) = 0$ characteristic poly
    - always $n$ $\lambda$, w/ multiplicity  eigenspace
    - each $\lambda_i$: $\Phi_i = N(\lambda_i I_n - A)$ subspace
  - Symmetric $A = A^T$
  - orthogonal $A^TA = AA^T = I$
  - dyad $ab^T$ (rank 1)
  - orthogonal-triangular decomp (QR)
    any square $A \to A = QR$ orthogonal, upper triangular
  - Singular value decomp (SVD)
    $\mathbb{R}^{m \times n}$ $A = U \tilde{\Sigma} V^T$
  - norm
    $f(A) \ge 0$
    $f(\alpha A) = |\alpha| f(A)$ "submultiplicativity"
    $f(A+B) \le f(A) + f(B)$
    $f(AB) \le f(A) f(B)$ if entrywise
  - Frobenius
    $\|A\|_F = \sqrt{\operatorname{trace} AA^T} = \sqrt{\sum \sum |a_{ij}|^2}$
    $= \sqrt{\sum \lambda_i(AA^T)}$ sum of eigen
    $\|Ax\|_2 \le \|A\|_F \|x\|_2$   $\sqrt{\sum \sigma_i^2(A)}$
    - submultiplicative
  - Operator norm
    $\|A\|_p = \max_{u \neq 0} \frac{\|Au\|_p}{\|u\|_p} = \max_{\|u\|=1} \|Au\|_p$
    $p = 1 \Rightarrow$ largest column sum
    $p = \infty \Rightarrow$ largest row sum
    $p = 2 \Rightarrow \sqrt{\lambda_{max}(A^TA)}$
- Matrix 2
  - Orthonormal basis (find w/ orthogonalization)
  - $\text{proj}_b a = \frac{a^Tb}{b^Tb} b$
  - Gram Schmidt - lin indep, lin. with subtraction
    - lin dep, if $q_i = 0$, skip
    - used for QR; length the same
    - lin dep, permute rows to make upper $\triangle$
  - Kernel trick
    - higher dim in data $K = X^TX$ kernel matrix
  - $\min L(X^Tw, y) + \lambda \|w\|_2^2$
    solution in span of $x_1, \dots, x_m$
- Kernel trick
  - nonlinear $\phi(x_i)$
  - $K_{ij} = \phi(x_i)^T \phi(x_j)$
  - kernel function $K(x, z) = \phi(x)^T \phi(z)$
  - ex. quadratic $(1 + x^Tz)^2$
    general poly $(1 + x^Tz)^p$
    $O(n)$ : instead of $O(n^p)$
  - ex. gaussian $K(x, z) = \exp\left(-\frac{\|x - z\|_2^2}{2\sigma^2}\right)$
- Symmetric Matrix
  - sample cov $C = \frac{1}{m} \sum_{i=1}^m (x^{(i)} - \hat{x})(x^{(i)} - \hat{x})^T$
    $C = \frac{1}{m} \tilde{X} \tilde{X}^T$ ; $\tilde{x}^{(i)}$ in columns, centered
  - $w^TCw$ = variance along $w$ (PCA max)
  - Hessian matrix
    $H_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}$ aka $\nabla^2 f(x)$ symmetric!
    quadratic $q(x) = \frac{1}{2} x^THx + c^Tx + d$
  - Spectral Theorem ($Q$ orthogonal, $\Lambda$ diag)
    - symmetric $\Rightarrow A = Q \Lambda Q^T$
      $= \sum_{i=1}^n \lambda_i v_i v_i^T$
    $\lambda_{min} \le \frac{x^TAx}{x^Tx} \le \lambda_{max}$
    Rayleigh quotient
  - matrix gain $\frac{\|Ax\|}{\|x\|}$
  - $\ell_2$ is spectral norm $= \sqrt{\lambda_{max}(A^TA)}$
    - along $v_1$
  - PSD - symmetric $x^TAx \ge 0$, $A \succeq 0$
  - PD - " , $x^TAx > 0$, $A \succ 0$
    iff invertible
    $A^TA \succeq 0$ and $AA^T \succeq 0$
    $A^TA \succ 0$ iff A full column rank
    $AA^T \succ 0$ iff A full row rank
  - Matrix square root
    $B = A^{1/2} \to A = B^2$
    $A = Q \Lambda Q^T$
    $B = Q \Lambda^{1/2} Q^T$
  - Ellipsoid
    - $\mathcal{E} = \{x : x^TP^{-1}x \le 1\}$, $P \succ 0$
    - $u_i$ axes, length $\sqrt{\lambda_i}$
    - Cholesky Decomp $P^{-1} = A^TA$
      $\mathcal{E} = \{x : \|Ax\|_2 \le 1\}$
  - Set of PSD is convex cone $S_+^n$
    - comparison order
  - Schur complement
    $M = \begin{pmatrix} A & X \\ X^T & B \end{pmatrix}$, $A, B$ symmetric, $B \succ 0$
    $S = A - XB^{-1}X^T$
    $\Rightarrow M \succeq 0 \iff S \succeq 0$
    works w/ $\succ$ too
- SVD
  - $A = U \tilde{\Sigma} V^T = \sum_{i=1}^r \sigma_i u_i v_i^T$
  - $r$ = rank$(A)$, $\sigma$ ordered descending $\ge 0$
  - $\sigma_i^2 = \lambda_i(A^TA) = \lambda_i(AA^T)$
  - $N(A)$ = last $n - r$ columns of V
  - $R(A)$ = first $r$ columns of U
  - $\|A\|_F^2 = \sum \sigma_i^2$
  - $\|A\|_2^2 = \sigma_1^2$
  - nuclear norm $\|A\|_* = \sum |\sigma_i|$
  - trace cyclic invariance
  - condition number $\kappa(A) = \frac{\sigma_1}{\sigma_n} = \|A\|_2 \cdot \|A^{-1}\|_2$
    $\uparrow \kappa \to$ close to singular, + sensitive to error
  - low rank approx
    $\min \|A - A_k\|_F^2 \Rightarrow$ first k of SVD! $\sum_{i=1}^k \sigma_i u_i v_i^T$
    rank$(A_k) = k$
  - total variance $\eta_k = \frac{\sigma_1^2 + \dots + \sigma_k^2}{\sigma_1^2 + \dots + \sigma_r^2}$
    error $e_k = 1 - \eta_k$
  - rank deficiency $\Rightarrow$ smaller $\sigma_n$!
  - PCA is SVD of centered matrix
    $\tilde{X}\tilde{X}^T$ eigenvalue, $\max z^T(\tilde{X}\tilde{X}^T)z$
  - Generalized low rank
    $\min_{L,R} \sum_{ij} \mathcal{L}(X_{ij}, l_i^T r_j) + \sum p_i(l_i) + \sum q_j(r_j)$ (loss, penalties)
    - Regularized PCA add $\gamma(\|L\|_F^2 + \|R\|_F^2)$
      - soln $\hat{\Sigma} = \max(0, \Sigma - \gamma)$  $L = U_k \hat{\Sigma}^{1/2}$
      - threshold $\sigma$  $R = V_k \hat{\Sigma}^{1/2}$
    - non-negative matrix factorization (NNMF), $L \ge 0$, $R \ge 0$
      - choose penalty
- Linear Equations
  - $Ax = y$
  - Vandermonde - poly $a_0 + a_1 t + a_2 t^2 \dots a_n t^n$
  - soln. iff rank$([A\ y])$ = rank$(A)$
    $S = \{x = \bar{x} + Nz\}$
    $A\bar{x} = y$  nullspace of A
    - unique soln if $N(A) = 0$
  - overdetermined $m > n$
  - underdetermined $m < n$
    - or plane
  - Square $m = n$
    - full rank + soln
  - Solve w/ SVD
    $Ax = y$, $A = U \tilde{\Sigma} V^T$
    $\tilde{\Sigma} \tilde{x} = \tilde{y}$
  - Solve w/ QR
    backsubstitution in triangular
- Least Squares + Variants
  - LS aka $Ax \approx y$
    $\min_x \|Ax - y\|_2^2$
    overdetermined $\to$ $x^* = (A^TA)^{-1}A^Ty$ full column rank
    - set of solns [illegible] $N(A)$
  - underdetermined
    $x^* = A^T(AA^T)^{-1}y$
  - OLS w/ QR
    $A^\dagger = (A^TA)^{-1}A^T = R^{-1}Q^T$
  - LS w/ equality
    $\min \|Ax - y\|_2^2$ s.t. $Cx = d$
    $\bar{x}$ s.t. $C\bar{x} = d$, $x = \bar{x} + Nz$
    $\min \|\tilde{A}z - \tilde{y}\|_2^2$
    $\tilde{A} = AN$  $\tilde{y} = y - A\bar{x}$
  - weighted LS
    $\min \|W(Ax - y)\|_2^2 = \|A_w x - y_w\|_2^2$
  - $\ell_2$-regularized LS
    add $\phi(x)$
    ex. $\phi(x) = \lambda \|x\|_2^2$ [illegible]
- Linear Programs
  - Half-space $a^Tx \le b$  $x_0 = \frac{b}{a^Ta} a$
  - Polyhedron $a_i^Tx \le b_i$:
    - intersection of half spaces
    - polytope = bounded
  - LP standard form - $f_i$ all affine
    $p^* = \min_x c^Tx + d$
    s.t. $A_{eq}x = b_{eq}$
    $Ax \le b$
    "inequality form"
  - Polyhedral Function
    - epigraph is polyhedron
      epi $f = \{(x, t) : C\begin{bmatrix} x \\ t \end{bmatrix} \le d\}$
    - pointwise max of function
      $f(x) = \max a_i^Tx + b_i$
      $a_i^Tx + b_i \le t$
    - $f(x) = \|x\|_\infty$ also [illegible] $\max(x, -x)$
    - $\ell_1$ too
    - can be cast as LP!!!
  - Relaxation
    - $|x| \le t \to$ $x \le t$, $-x \le t$
- Convex Quadratic Program
  $f_0(x) = \frac{1}{2} x^THx + c^Tx + d$
  H is symmetric
  $\nabla f = \frac{1}{2}(H + H^T)x + c$
  - convex if $H \succeq 0$ ($\lambda \ge 0$)
  - QP
    $\min f(x)$
    s.t. $Ax \le b$, $Cx = d$
  - unconstrained? $x^* = -H^{-1}c$
    $-p^* = -\frac{1}{2} c^T H^{-1} c$
  - linear eq. [illegible]
    $x = \bar{x} + Nz$, [illegible]
  - low cardinality ($\ell_1$ of $\ell_0$)
    - sparsity
    - hard use approximation
  - Piecewise Constant Fitting
  - BPDN/LASSO
    $\min \|Ax - y\|_2^2 + \lambda \|x\|_1$
    QP w/ slack variable [illegible]
- Second Order Cone Program
  - superset of LP, QP
  - Second Order Cone (SOC)
    $\sqrt{x_1^2 + x_2^2} \le t$
  $\mathbb{R}^{n+1} \to K_n = \{(x, t) : \|x\|_2 \le t\}$
  - rotated SOC $\mathbb{R}^{n+2}$
    $K_n^r = \{(x, y, z) : x^Tx \le 2yz, y \ge 0, z \ge 0\}$
    equiv $\|(x, (y - z)/\sqrt{2})\|$
    $t = (y + z)/\sqrt{2}$
    - hyperbolic constraint
  - standard SOC
    $\|Ax + b\|_2 \le c^Tx + d$
  - QP $\to$ SOCP
    $\min y^TQx + c^Tx$  $Q \succeq 0$
    $a_i^Tx \le b_i$
    $\min c^Tx + y$  $y \ge x^TQx$
    $\left\| \begin{matrix} 2Q^{1/2}x \\ y - 1 \end{matrix} \right\|_2 \le y + 1$
    $a_i^Tx \le b_i$
  - QCQP $\to$ SOCP
    $\min x^TQ_0x + a_0^Tx$
    s.t. $x^TQ_ix + a_i^Tx \le b_i$
    $\min a_0^Tx + t$
    $\left\| \begin{matrix} 2Q_0^{1/2}x \\ t - 1 \end{matrix} \right\|_2 \le t + 1$
    $\left\| \begin{matrix} 2Q_i^{1/2}x \\ b_i - a_i^Tx - 1 \end{matrix} \right\|_2 \le b_i - a_i^Tx + 1$
- Robust Optimization
  - uncertainty affects solutions
  - sacrifice performance for robust
  - constraint
    $\min f_0(x)$, $f_i(x) \le 0$
    "nominal case"
  $\min_x \max_u f_0(x, u)$
    $f_i(x, u) \le 0$  constraint
  - add uncertainty $u \in \mathcal{U}$
  - single inequality w/ uncertain coefficients
    $\forall a \in \mathcal{U}$, $a^Tx \le b$
    - $b \ge \phi_{\mathcal{U}}(x) := \max_{a \in \mathcal{U}} a^Tx$
    - box $\mathcal{U} = \{a : \|a - \hat{a}\|_\infty \le \rho\}$
      $= \{\hat{a} + \rho u : \|u\|_\infty \le 1\}$
      $\max_{a \in \mathcal{U}} a^Tx = \hat{a}^Tx + \rho \max u^Tx$
      $= \hat{a}^Tx + \rho \|x\|_1$
    - sphere / ellipse $\|u\|_2 \le 1$
      $\max = \hat{a}^Tx + \rho \|x\|_2$
    - ellipsoid $P = R^TR$
      $\max = \dots + \|R^Tx\|_2$
  - Chance LP
    - probability satisfy > some
  - Robust LS
    - here $\|A - \hat{A}\| \le \rho$
    $\min_x \max_{\|\Delta\| \le \rho} \|(\hat{A} + \Delta)x - y\|_2$
    $\|\hat{A}x - y\|_2 + \rho \|x\|_2$
    $\Delta = \frac{\rho}{\|\hat{A}x - y\|_2 \|x\|_2} (\hat{A}x - y)x^T$
    - becomes regularized LS
- Extra
  - $\|Rx\|_2^2 \le c^Tx$ can be SOCP
  - relax QP $\to$ QCQP

Matthew Tran
EE127
Final

- Convexity
  - Set of points in $\mathbb{R}^n$, $P$
    - linear hull - all linear combos
    - affine hull - coeff sum to 1 (aff $P$)
      - smallest affine set w/ $P \subseteq$ aff $P$
    - Convex combo - $\lambda \geq 0$, $\sum \lambda = 1$
      - convex hull - set all ↑
      - conic hull - ↑ but no $\sum \lambda = 1$
  - Convex - has line segment between 2 pts: $\lambda x_1 + (1-\lambda)x_2 \in C$, $\lambda \in (0,1)$
  - cone - $x \in C \Rightarrow \alpha x \in C$
    - conic hull of convex set [illegible] convex cone
  - intersection of convex sets is convex
  - affine transformation of convex is convex $(Ax+b)$
  - convex function ∪ → $+\infty$ outside dom $f$
    - dom $f$ is convex
    - $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y)$
  - concave - $-f$ is convex
  - epigraph - pts. above function
    $f: \mathbb{R}^n \to \mathbb{R} \cup \{-\infty, +\infty\}$
    epi $f = \{(x,t) : x \in \text{dom } f,\ t \in \mathbb{R},\ f(x) \leq t\}$
  - Sublevel Set $S_\alpha = \{x \in \mathbb{R}^n : f(x) \leq \alpha\}$
    $f$ convex $\Rightarrow S_\alpha$ convex (not ⇐ usually)
  - $f(x) = \sum \alpha_i f_i(x)$, $\alpha_i \geq 0 \Rightarrow$ preserves convex over $f_i$
  - First Order Conditions
    $f$ convex $\iff \forall x, y \in \text{dom } f$, $f(y) \geq f(x) + \nabla f(x)^T (y - x)$
    $f$ is bounded below by any hyperplane tangent
  - Second Order Conditions
    Hessian $\nabla^2 f$ is PSD $\nabla^2 f \succeq 0$
  - Pointwise Maximum
    $f(x) = \max_i f_i(x)$ is convex
    used in SOCP $\|Ax+b\|_2 - (c^T x + d)$ is convex
- Convex Problems
  LP ⊆ QP ⊆ QCQP ⊆ SOCP ⊆ CP
  - Standard form
    $p^* = \min_{x \in \mathbb{R}^n} f_0(x)$ s.t.
    $f_i(x) \leq 0$ ← convex
    $Ax = b$ ← affine
    - implied $x \in \text{dom } f_0 \cap \text{dom } f_i$
  - implicit constraints
  - SOCP can be all convex probs
  - equivalent form
    $p^* = \min_{x \in X} f_0(x)$ ← convex
    - unconstrained if $X = \mathbb{R}^n$
    - infeasible $X$ empty (by $p^* = \infty$)
    - unbounded $p^* = -\infty$
  - possible bounded but no feasible pt. (ex <)
  - $f(x^*) < \infty$ (not ≥) ⇒ [illegible]
    - can remove [illegible] convex
  - optimal set $X_{opt}$ ← always convex
  - convex ⇒ local opt = global opt
  - Transforms
    - monotone objective transformation
      - continuous and strictly increasing
      - ex. log
    - slack variables - good for linear objective [illegible]
    - equality ⇒ inequality (= ⇒ ≤)
      $p^* = \min f_0(x)$ s.t. $h(x) = u$; $g^* = \min f_0(x)$ s.t. $h(x) \leq u$
      - $p^* = g^*$ if
        - $f_0$ nonincreasing over $X$ (decrease $f_0$ more)
        - $h$ nondecreasing over $X$
      - convex if $f_0$ is strictly monotone
- Hidden Convexity
  - "also" nonconvex problems "exact" convex optimal

- Weak Duality
  - Optimization Problem for
    $p^* = \min_x f_0(x)$
    maybe nonconvex ⇒ $f_i(x) \leq 0$
    $h_i(x) = 0$
    "primal"
  - Lagrangian
    $\mathcal{L}(x, \lambda, \nu) = f_0(x) + \sum \lambda_i f_i(x) + \sum \nu_i h_i(x)$
    $\lambda \geq 0$, $\nu$ are Lagrange multipliers or dual variables
    $p^* = \min_x \max_{\lambda \geq 0, \nu} \mathcal{L}(x, \lambda, \nu)$ ← over entire domain
  - Minimax Inequality
    $p^* = \min_x \max_y F(x,y) \geq \max_y \min_x F(x,y) = d^*$
  - Weak duality $p^* \geq d^*$
  - ☆ dual is convex! even if primal is nonconvex
  - dual switches [illegible] # of variables
- Strong Duality
  - convex problem
    $p^* = \min_x f_0(x)$
    convex ⇒ $f_i(x) \leq 0$
    $Ax = b$
    ⇓
    $\mathcal{L}(x, \lambda, \nu) = f_0(x) + \sum \lambda_i f_i(x) + \nu^T (Ax - b)$
  - Relative Interior
    convex $D = \{x \in \mathbb{R}^n : Cx \leq d,\ Fx = g\}$
    relint $D = \{x \in \mathbb{R}^n : Cx < d,\ Fx = g\}$
  - Strictly feasible
    $\exists x_0 \in$ relint $D$ st. $Ax_0 = b$, $f_i(x_0) < 0$
  - ☆ Slater's Condition for Convex Programs
    strictly feasible $\Rightarrow p^* = d^*$
    - basically nonempty $p, d$ attainable
  - Primalization
    - w/ strong duality $p^* = d^*$ and have $\lambda^*, \nu^*$, get $x^*$?
    - find $x^* = \min \mathcal{L}(x, \lambda^*, \nu^*)$ to get the min $p^*$
      - if unique, then check feasibility
  - unconstrained problem (Lagrangian)
    - dual is constant (no vars) same $p^*$ [illegible]
    - add variables! (slack vars) [illegible]
  - ☆ Sion's minimax theorem
    Let $X \subseteq \mathbb{R}^n$ be convex, $Y \subseteq \mathbb{R}^m$ be compact set
    $F(x,y)$ st. $\forall y$ $F(x,y)$ convex cont. over $X$ ∪
    $\forall x$ $F(x,y)$ concave cont. over $Y$ ∩
    $\max_y \min_x F(x,y) = \min_x \max_y F(x,y)$
    ↑ like saddle point
  - Safe Feature Elimination (SAFE)
    - $f$ convex, can remove inactive [illegible]

$C = \frac{1}{m} X X^T$

- Cov $C$ PSD iff $X$ is centered
- $x^T C^{-1} x \leq 1$ is bounded ellipsoid
- SVD $X = USV^T$
  $\det(U) = \pm 1$, not unique
- Schur complement
  $A \in S^n$, $B \in S^m$, $X \in \mathbb{R}^{n \times m}$, $B \succ 0$
  $M = \begin{bmatrix} A & X \\ X^T & B \end{bmatrix}$ $S = A - XB^{-1}X^T$
  $M \succ 0 \iff S \succ 0$
- QP $A \succeq 0$ ∪, unique if $A \succ 0$
- projection $P^2 = P$

- $\mathbb{R}_+ \Leftrightarrow x \geq 0$, $\mathbb{R}_{++} \Leftrightarrow x > 0$
- convex/concave test
  - midpoint check $\forall A$ (convex, concave)
  - $f''(x)$
  - composition $f$ convex ⇒ convex [illegible]
  - [illegible] ⇒ not convex by midpoint
  - $x \log x$ strictly convex [illegible]
- $x_1 x_2$ not convex
- $\sigma_{min}(A) = \min_{v \neq 0} \frac{\|Av\|_2}{\|v\|_2}$
- change of variables:
- $A \succ 0$ $A = A^{1/2} A^{1/2}$
- invertible $\iff$ no $\lambda = 0$
- commute $PQ = QP$
- Cauchy $|x^T y| \leq \|x\|_2 \|y\|_2$

- Algorithms
  - 1D, Finite interval
    - $f$ easy ⇒ gridding
    - $f$ more diff ⇒ better
      - extend to [illegible]
  - 1D, Bisection for convex
    - update interval to middle (1 bit binary search w/ $f'(x)$)
    - $O(\log(1/\epsilon))$ for $\epsilon$-optimal
    - $x^*$ = endpoint ⇒ expand at each [illegible]
  - problem decomposition
    - split into easy pieces
    - Fenchel conjugate $f_i^*$
      $f_i^*(a) = \max_\xi x\xi - f_i(\xi)$
      - convex
  - Coordinate descent
    - good for "box" problems $\|x\|_\infty \leq 1$
    - also "product" $\min f(x) + \lambda \|x\|_1$
    - idea:
      - optimize 1 variable at time in 1D.
      - proven global min if $f$ strictly convex
    - "independent" constraints
  - Gauss-Seidel method (Block Coordinate minimization)
    $x_i^{(k+1)} = \arg\min_{x_i \in X_i} f(x_{i-1}^{(k+1)}, x_i, x_{i+1}^{(k)}, \ldots)$
    - idea go 1 to $k$, get $x_i^{(k+1)}$ one by one using old $x^{(k)}$ for [illegible]
    - Assume $f$ convex, cont. diff on $X$, strictly minimizing [illegible] ⇒ converges to optimal
    - Usually fails if non-smooth, except [illegible] smooth
  - Basic Gradient method $x^{(k+1)} = x^{(k)} - s_k \nabla f_0(x^{(k)})$
    - slow, [illegible]
  - Stochastic Gradient Descent (SGD)
    - min of sums of [illegible]
    - update using each point or batch "minibatch"
      $\nabla f_i(w) \approx L'(w^T x_i) x_i$
    - converges, but slow
  - Projected Gradient
    $\min_{x \in X} f(x)$ where $X$ is "simple" if can project any point $x$ on $X$
    $x^{(k+1)} = \pi_X(x^{(k)} - s_k \nabla f_0(x^{(k)}))$
    - dual might help
  - Newton's method for unconstrained
    $\min_x f(x)$, have $\nabla f(x)$, $\nabla^2 f(x)$
    $x^{(k+1)} = \nabla^2 f(x)^{-1} \nabla f(x) + x^{(k)}$
    [illegible] $x^{(k+1)} = x^{(k)} + \frac{f'(x^{(k)})}{f''(x^{(k)})}$
    [illegible]
  - Interior point for constrained prob
    $\min_x f(x)$, $f_i(x) \leq 0$
    ⇓
    cvx! → $\min_x f(x) - \mu \sum \log(-f_i(x))$, $t > 0$
    ↑ now unconstrained
    - iterate drives "interior" to feasible
    - [illegible] ⇒ $\epsilon$-optimal
    - change [illegible]
    - fast converge, [illegible] steps

- Application to ML
  - Supervised Learning
    - has regression, regression or classification
    - goal for
      $\min_w L(X^T w, y) + \lambda p(w)$
      [illegible]
    - Loss function
      - squared $L(z,y) = \|z - y\|_2^2$
      - $\ell_1$ norm $L(z,y) = \|z - y\|_1$ outliers
      - hinge $L(z,y) = \sum \max(0, 1 - y_i z_i)$ SVM
      - logistic $L(z,y) = -\sum \log(1 + e^{-y_i z_i})$ log regression
    - Linear Least Squares
      $\min_w \|X^T w - y\|_2^2 + \lambda \|w\|_2^2$ ← ridge regression
      $+ \lambda \|w\|_2^2 + \mu \|w\|_1$ LASSO + elastic net
      $\min_w \|X^T w - y\|_2^2 + \lambda \sum \|w_i\|_2$ group LASSO
    - Linear Classification
      - $y \in \{-1, 1\}$ rule $\hat{y} = \text{sign}(w^T x + b)$
      - separable
      - hyperplane $H = \{w : w^T x = c\}$
        dist to origin $|c| / \|w\|_2$, $x_0 = \frac{cw}{w^T w}$
      - $y_i(w^T x_i + b) \geq 1$
      - 2 hyperplanes $d = 2 / \|w\|_2$
      - $L(w,b) = \sum \max(0, 1 - y_i(w^T x_i + b))$
      - dual, smallest dist between +1 and −1
      - maximum likelihood
        $\ell(w,b) = \prod_i \frac{1}{1 + e^{-y_i(w^T x_i + b)}}$
        max [illegible] $= -\sum \log(1 + e^{-y_i(w^T x_i + b)})$
      - [illegible] $\min_w L(Z^T w, y)$ $Z \sim \mathcal{N}(0, [illegible])$
      - [illegible] high dim,
        - [illegible] $\ell_1$-norm regularization
  - Unsupervised Learning
    [illegible]
    - PCA not [illegible] to noise at all, [illegible]
    - Low rank model
      $\min_{L,R} \|X - LR^T\|_F$ convex penalty
      $\min_{L,R} \sum L(X_{ij}, l_i^T r_j) + \sum p_i(l_i) + \sum q_j(r_j)$
    - Robust PCA
      $X = LR^T + S$ ← sparse ⇒ $\min_{L,R} \|X - LR^T\|_1$
    - NNMF (Non-negative Matrix Factorization)
      $X = LR^T$, $L \geq 0$, $R \geq 0$ (data ≥ 0)
    - Sparse PCA
    - Matrix Completion
      - [illegible] low rank
        $\min_{L,R,X \in \mathcal{X}} \|X - LR^T\|_F^2 + \gamma(\|L\|_F^2 + \|R\|_F^2)$
        alternate $L$ and $R$
    - Sparse Covariance Selection
      - [illegible] graph
      - Gaussian assumption
        $p(x) = \frac{1}{(2\pi \det \Sigma)^{n/2}} e^{-\frac{1}{2} x^T \Sigma^{-1} x}$
      - conditional independence
        $p(x) = p_i(x_i) p_j(x_j)$ [illegible] all other
        iff $(\Sigma^{-1})_{ij} = 0$