Classics

Classes	Solutions
Projection and Mechanics	
$\min_{\alpha\in\mathbb{R}} \|x-\alpha u\|_2^2$	$\alpha = \frac{x^T u}{\|u\|_2^2} = \frac{u^T x}{\|u\|_2^2}$
$\min_{x\in\mathbb{R}} \|v - x\mathbb{1}\|_2^2$	$x = \frac{1}{n}\sum_{i=1}^{n} v_i = \text{average}$
$\min_{x\in\mathbb{R}^m} \|X - x\mathbb{1}^T\|_F$	$x = \frac{1}{n} X\mathbb{1} = \text{column average}$
$\min_{x\in\mathbb{R}} \|v - x\mathbb{1}\|_1$	$x = \text{median}$
$\min_{w\in\mathbb{R}^m} L(A^T w) + \lambda\|w\|_2^2, A \in \mathbb{R}^{m\times n}, \lambda \geq 0$	$\min_{v\in\mathbb{R}^n} L(A^T A v) + \lambda\|Av\|_2^2$
Least Squares and Variants	
Ordinary Least Squares $A \in \mathbb{R}^{m\times n}, x \in \mathbb{R}^n, y \in \mathbb{R}^m$	
$\min_x \|Ax - y\|_2$	$x_{opt} = A^{pi} y + N(A)$ $x^* = A^{pi} y \in R(A^T) = N(A)^\perp$ (least norm) If A full column rank, solution is unique: $x^* = (A^T A)^{-1} A^T y, \ \ Rx^* = Q^T y$
$\arg\min_\beta \sum_{i=1}^{n} (y_i - \beta x_i)^2$	$\left(\sum_{i=1}^{n} x_i^2\right)\beta^* = \sum_{i=1}^{n} x_i y_i$
Linearly Constrained Least Squares / Perturbations to Feasibility $C \in \mathbb{R}^{p\times n}, d \in \mathbb{R}^p$	
$\min_x \|Ax - y\|_2^2 \quad \text{s.t. } Cx = d$	$x = C^{pi} d + Nz$ where columns of N form a basis for $N(C)$
$\arg\min_x \|x\|_2^2 \quad \text{s.t. } Ax = y$	$x = A^{pi} y$ provided $y \in R(A)$, else no solution exists
$\min_{\delta y} \|\delta y\|_2^2 \quad \text{s.t. } y + \delta y \in R(A)$	$\delta y = y - AA^{pi} y = (\mathbb{I} - AA^{pi}) y$
Weighted Least Square	
$\min_x \sum_{i=1}^{m} w_i^2 \left\|a_i^T x - y\right\|^2$	$\min_x \|W(Ax - y)\|_2^2$ where $W = \text{diag}(w_1, \ldots, w_m)$
$\min_x \|Ax - y\|_2^2 + x^T W x, W \succ 0$	$\min_x \left\| \begin{bmatrix} A \\ W^{\frac{1}{2}} \end{bmatrix} x - \begin{bmatrix} y \\ 0 \end{bmatrix} \right\|_2^2$ $x^* = (A^T A + W)^{-1} A^T y$
Regularized Least Squares / Ridge Regression	
$\min_x \|Ax - y\|_2^2 + \lambda^2 \|x - c\|_2^2, \qquad \lambda > 0$	$\min_x \left\| \begin{bmatrix} A \\ \lambda\mathbb{I} \end{bmatrix} x - \begin{bmatrix} y \\ \lambda c \end{bmatrix} \right\|_2^2$ always full column rank due to $\lambda\mathbb{I}$ $x^* = (A^T A + \lambda^2 \mathbb{I})^{-1} (A^T y + \lambda^2 c)$
$\arg\min_x \|Ax - b\|_2^2 + \|\Gamma x\|_2^2$	$\arg\min_x \left\| \begin{bmatrix} A \\ \Gamma \end{bmatrix} x - \begin{bmatrix} b \\ 0 \end{bmatrix} \right\|_2^2$ Assuming Γ full rank: $x^* = (A^T A + \Gamma^T \Gamma)^{-1} A^T b$
Kernel Least Squares / Kernel Trick	
$\min_w \|A^T w - y\|_2^2 + \lambda\|w\|_2^2$	$\min_v \|A^T A v - y\|_2^2 + \lambda\|Av\|_2^2$ i.e. optimal w lies in $R(A)$ If $n \gg m$, dramatic reduction in problem size $\min_v \left\| \begin{bmatrix} A^T A \\ \sqrt{\lambda} A \end{bmatrix} v - \begin{bmatrix} y \\ 0 \end{bmatrix} \right\|_2^2$ $v = (K^2 + \lambda K)^{-1} K y, K = A^T A, w = Av$

Variational Characterization of Eigenvalues / Rayleigh Quotient / Maximum Gain	
$\max_{x:\|x\|_2\le 1}\|Ax\|_2 = \max_{x:\|x\|_2=1}\frac{\|Ax\|_2}{\|x\|_2}$	$\sqrt{\lambda_{\max}(A^TA)}$
Maximum / Minimum Variance Direction $A = A^T \in \mathbb{S}^n$	
$\max_{x\neq 0}\frac{x^TAx}{x^Tx} = \max_{\|x\|_2\le 1} x^TAx$ $\arg\max_{\|x\|_2\le 1} x^TAx$	$\lambda_1(A)$ u_1
$\min_{x\neq 0}\frac{x^TAx}{x^Tx} = \min_{\|x\|_2\le 1} x^TAx$ $\arg\min_{\|x\|_2\le 1} x^TAx$	$\lambda_n(A)$ u_n
$\max_{\mathcal{V}:\dim\mathcal{V}=k}\min_{x\in\mathcal{V}}\frac{x^TAx}{x^Tx}$ $\min_{\mathcal{V}:\dim\mathcal{V}=k}\max_{x\in\mathcal{V}}\frac{x^TAx}{x^Tx}$	$\lambda_k(A)$ $\lambda_{n-k+1}(A)$
$\min_{X':\mathrm{rank}(X')\le 1}\|X-X'\|_F = \min_{u,v}\|X-uv^T\|_F$ $\min_{u\in\mathbb{R}^m:\|u\|_2=1}\sum_{i=1}^n \min_{\alpha_i\in\mathbb{R}}\|x_i-\alpha_i u\|_2^2$ $\max_{u\in\mathbb{R}^m:u^Tu=1} u^TCu,\ C=\frac{1}{n}XX^T$	Assuming X is centered. $\alpha_i = x_i^Tu$ $X' = u[\alpha_1 \cdots \alpha_n]$ $\lambda_{\max}(C)$
$\min_{U\in\mathbb{R}^{n\times n},U\neq 0}\frac{tr(UAU^T)}{\lambda_{\max}(UBU^T)}, A, B \succ 0$	$\lambda_{min}\left(B^{-\frac{1}{2}}AB^{-\frac{1}{2}}\right)$
$\min_{\lambda:\lambda\mathbb{I}-C\succcurlyeq 0}\lambda\mathbb{I}-C$	$\lambda = \lambda_{\max}(C)$
$\min_{x,y:\|x\|_2,\|y\|_2\le 1} x^TAy$	$-\sigma_{\max}(A)$
$\max_{x,y:\|x\|_2,\|y\|_2\le 1} x^TAy$	$\sigma_{\max}(A)$
Singular Value Decomposition $A\in\mathbb{R}^{m\times n}$	
$\mathrm{argmax}_{v:\|v\|_2\le 1}\|Av\|_2$	v_1 $u_1 = \frac{Av_1}{\sigma_1}$
$\|A\|_2 = \max_{v:\|v\|_2\le 1}\|Av\|_2$	$\sigma_1(A) = \|Av_1\|_2$
$\min_{V\subset\mathbb{R}^n, \dim V=k}\sum_{i=1}^m\|a_i-\Pi_V(a_i)\|_2^2$	$V = \mathrm{Span}\{v_1,\dots,v_k\}$ $\sigma_{k+1}^2+\cdots+\sigma_r^2$
$\min_{x,y}\|A-xy^T\|_F$	$\sqrt{\sigma_2^2+\cdots+\sigma_r^2}$
Rank k Approximation / Distance to Rank Deficiency	
$\arg\min_{A_k\in\mathbb{R}^{m\times n}, \mathrm{rank}(A_k)=k}\|A-A_k\|_F$ $\min_{A_k\in\mathbb{R}^{m\times n}, \mathrm{rank}(A_k)=k}\|A-A_k\|_F$	$A_k = \sum_{i=1}^k\sigma_iu_iv_i^T$ $\sqrt{\sigma_{k+1}^2+\cdots+\sigma_r^2}$
$\min_{V_k\subset\mathbb{R}^m,\dim V_k=k}\sum\|a_i-\Pi_k(a_i)\|_2^2$ a_i: ith row of A	$\sqrt{\sigma_{k+1}^2+\cdots+\sigma_r^2}$
$\min_{L,R}\|X^c-LR^T\|_F, L\in\mathbb{R}^{m\times k}, R\in\mathbb{R}^{k\times n}$ $\min_{L,R}\|X^c-LR^T\|_F, L\in\mathbb{R}^{m\times k}, R\in\mathbb{R}^{k\times n}, R^T\mathbb{1}=0$	R consists of the top k right singular vectors of the SVD of X^c. $\mathbb{1}\in N(X^c)\subset N(R^T)$
Linear Programming (LP)	

Unconstrained LP $A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m, c \in \mathbb{R}^n$	
$\min_x c^T x \quad \text{s.t.} \quad Ax \le b$	$p^* = \begin{cases} 0, & c = 0 \\ -\infty, & c \ne 0 \end{cases}$
$\min_x c^T x + \lambda \lVert x \rVert_1 \quad \text{s.t.} \quad Ax \le b, \lambda > 0$	$\min_{x,u} c^T x + \lambda \sum_{i=1}^{n} u_i \quad \text{s.t.} \quad Ax \le b, u_i \ge \lvert x_i \rvert$
l_∞ Regression / Minimum Robust System $A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m$	
$\min_x \lVert Ax - b \rVert_\infty$ $\min_x \max_{1\le i\le m} \lvert a_i^T x - b_i \rvert$ $\min_{x,t} t \ \text{s.t.} \ \lvert a_i^T x - b_i \rvert < t \ \forall i$	$\min_{x,t} \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ t \end{bmatrix}$ subject to $\begin{bmatrix} A & -\mathbb{1} \\ -A & -\mathbb{1} \end{bmatrix} \begin{bmatrix} x \\ t \end{bmatrix} \le \begin{bmatrix} b \\ -b \end{bmatrix}$
l_1 Regression $A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m$	
$\min_x \lVert Ax - b \rVert_1 = \min_x \sum_{i=1}^{m} \lvert a_i^T x - b_i \rvert$ $\min_u \mathbb{1}^T u \quad \text{s.t.} \quad \lvert a_i^T x - b_i \rvert < u_i \ \ \forall i$	$\min_{x,u} \begin{bmatrix} 0 & \mathbb{1}^T \end{bmatrix} \begin{bmatrix} x \\ u \end{bmatrix}$ subject to $\begin{bmatrix} A & -\mathbb{I} \\ -A & -\mathbb{I} \end{bmatrix} \begin{bmatrix} x \\ u \end{bmatrix} \le \begin{bmatrix} b \\ -b \end{bmatrix}$
Assignment Problem (M_{ij} = time for worker j to complete task i)	
$\min_{X\in\{0,1\}^{m\times n}} tr(X^T M) = \min_{X\in\{0,1\}^{m\times n}} \sum_{i,j}^{m,n} X_{ij} M_{ij}$ subject to $x\mathbb{1} = \mathbb{1}$, $x^T \mathbb{1} \le \mathbb{1}$	$\min_{0\le X_{ij}\le 1} tr(X^T M) = \min_{0\le X_{ij}\le 1} \sum_{i,j}^{m,n} X_{ij} M_{ij}$ subject to $x\mathbb{1} = \mathbb{1}$, $x^T \mathbb{1} \le \mathbb{1}$
Max Flow Problem	
$\max_{f,x} f$ subject to $Ax = \begin{bmatrix} -f & 0 & \cdots & 0 & f \end{bmatrix}^T$	• Conservation of flow: $(Ax)_j = 0$ • x_i: flow along link i • $(Ax)_j$: net flow out of vertex j
Boolean Linear Program (Relaxation)	
$\min_x c^T x$ subject $Ax \le b$, $x_i \in \{0,1\}$	$\min_x c^T x$ subject $Ax \le b$, $x_i \in [0,1]$
Cardinality Minimization Trick (Relaxation)	
$\min_x \text{card}(x)$	• Replace $\text{card}(x)$ with $\lVert x \rVert_1$ • Replace $\lVert x \rVert_2$ with $\frac{\lVert x \rVert_1}{\sqrt{\text{card}(x)}}$ • $\lVert x \rVert_1 \le \lVert x \rVert_2 \sqrt{\text{card}(x)}$
Constraint Manipulation	
$u_i \ge \lvert x_i \rvert$	$u_i \ge -x_i, \qquad u_i \ge x_i$
$\max_{i=1,\dots,m} \lvert a_i^T x - b_i \rvert \le t$	$a_i^T x - b_i \le t, \qquad -a_i^T x + b_i \le t$
Quadratic Programming (QP)	
Unconstrained QP	
$\min_x \frac{1}{2} x^T H x + c^T x \quad \text{s.t.} \quad Ax \le b, Cx = d$ $x \in \mathbb{R}^n, H \in \mathbb{S}$	• $H \nsucceq 0$ i.e. H has negative eigenvalues $p^* = -\infty$ • $H \succeq 0$ invertible $p^* = -\frac{1}{2} c^T H^{-1} c, \ \ x^* = -H^{-1} c$ • $H \succeq 0$ not invertible, $c \in R(H)$ $p^* = -\frac{1}{2} c^T H^{pi} C, \qquad x^* = -H^{pi} c + N(H)$ • $H \succeq 0$ not invertible, $c \notin R(H)$ $p^* = -\infty$
Least Squares	
$\min_x \lVert Ax - y \rVert_2^2$	$\min_x \frac{1}{2} x^T (2A^T A) x - 2y^T A x$

	$H = 2A^TA, c = -2A^Ty$
QP with Equality Constraints	
$\min_x \frac{1}{2}x^THx + c^Tx \text{ s.t. } Ax = b$	$H' = N^THN$ $c' = N^Tc + N^TH^Tx_0$
LASSO / l_1 Regularized Least Squares	
$\min_x \\|Ax-y\\|_2^2 + \lambda\\|x\\|_1$	$\min_{x,t} \\|Ax-y\\|_2^2 + \lambda\sum_{i=1}^n t_i \quad \text{s.t.} \quad t \ge x, t \ge -x$
$\min_{x\in\mathbb{R}} \frac{1}{2}\\|ax-y\\|_2^2 + \lambda\|x\|$	$x^* = \begin{cases} 0, & \|a^Ty\| \le \lambda \\ \frac{a^Ty}{\\|a\\|_2^2} - \frac{\lambda}{\\|a\\|_2^2}\text{sign}(a^Ty), & \|a^Ty\| > \lambda \end{cases}$
Cardinality Minimization Constraints (Relaxed)	
$\min_x \\|Ax-y\\|_2^2 \quad \text{s.t.} \quad \text{card}(x) \le k$	$\min_x \\|Ax-y\\|_2^2 \quad \text{s.t.} \quad \\|x\\|_1 \le k\\|x\\|_\infty$ $\min_x \\|Ax-y\\|_2^2 + \lambda\\|x\\|_1$
Piecewise Constant Function Fitting	
$D = \begin{bmatrix} -1 & 1 & 0 & 0 & \dots & 0 \\ 0 & -1 & 1 & 0 & \dots & 0 \\ 0 & 0 & -1 & 1 & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & -1 & 1 \end{bmatrix}$ $\min_x \\|x-y\\|_2^2 \quad \text{s.t.} \quad \text{card}(Dx) \le k$	$\min_x \\|x-y\\|_2^2 + \lambda\\|Dx\\|_1$
Second-order Cone Programming (SOCP)	
$\min_x c^Tx \quad \text{s.t.} \\|A_ix+b_i\\|_2 \le c_i^Tx + d, i = 1, \dots, m$	$A_i \in \mathbb{R}^{m_i\times n}, b_i \in \mathbb{R}^m, d_i \in \mathbb{R}, c \in \mathbb{R}^n, i = 1, \dots, m$
Reduction of QP	
$\min_x \frac{1}{2}x^THx + c^Tx \quad \text{s.t.} \quad Ax \le b$	$\min_{x,y} y + c^Tx$ subject to $\frac{1}{2}x^THx \le y, Ax \le b$
Quadratically Constrained QP (QCQP)	
$\min_x \frac{1}{2}x^THx + c^Tx \quad \text{s.t.} \quad x^TQ_ix + a_i^Tx \le b_i$ $Q_i \in \mathbb{S}, Q_i \succcurlyeq 0$ for $i = 1, \dots, m$	$\min_{x,u} u$ subject to $x^TQ_ix + a_i^Tx \le b_i, \frac{1}{2}x^THx + c^Tx \le u$
Reciprocals	
$\min_x \sum_{i=1}^n h_ix_i + \frac{c_i}{x_i}$ subject to $0 \le x$	$\min_{x>0} \sum_{i=1}^n h_ix_i + c_iy_i$ subject to $0 \le x, \frac{1}{x_i} \le y_i \Rightarrow 1 \le x_iy_i$
Rational Powers	
$\min_w \\|X^Tw-y\\|_2 + \lambda\sum_{i=1}^n \|w_i\|^{\frac{3}{2}}$	$\min_{w,u,v,t} \\|X^Tw-y\\|_2 + \lambda\sum_{i=1}^n t_i$ subject to $t \ge 0, u_i \ge \|w_i\|, v_it_i \ge u_i^2, u_i \ge v_i^2$
Facility Locations	
$\min_x \sum_{i=1}^n \\|A_ix-y_i\\|_2$	$\min_{x,t} \sum_{i=1}^n t_i$ subject to $\\|A_ix-y_i\\|_2 \le t_i$ for $i = 1, \dots, m$
$\min_x \max_i \\|x-y_i\\|_2$	$\min_{x,t} t$ subject to $\\|A_ix-y_i\\|_2 \le t_i$ for $i = 1, \dots, m$

$\min_x \sum_{i=1}^{m} \|A_i x - b_i\|_2 + \lambda\|x\|_1 + \mu\|x\|_\infty$	$\min_{x,y,u,t} \sum_{i=1}^{m} y_i + \lambda \sum_{i=1}^{n} u_i + \mu t$ subject to $\|A_i x - b_i\|_2 \le y_i, \|x_i\| \le u_i, \|x_i\| \le t$
Constraint Manipulation	
$Cx \le r$	$A_i = 0, b_i = 0, c_i = -(i\text{th row of } C), d_i = r_i$
$x^T Q x + c^T x \le t$ $Q \succcurlyeq 0$	$\left\\|\begin{bmatrix} \sqrt{2} Q^{\frac{1}{2}} \\ -c^T \end{bmatrix} x + \begin{bmatrix} 0 \\ t - \frac{1}{2} \end{bmatrix}\right\\|_2 \le t - c^T x + \frac{1}{2}$
$\|Rx\|_2^2 \le r^T x$	$\left\\|\begin{bmatrix} \sqrt{2} R \\ r^T \end{bmatrix} x - \begin{bmatrix} 0 \\ 1 \\ \frac{1}{2} \end{bmatrix}\right\\|_2 \le r^T x + \frac{1}{2}$
$\|x\|_2^2 \le 2yz$	$\left\\|\begin{bmatrix} x \\ \frac{1}{\sqrt{2}}(y - z) \end{bmatrix}\right\\|_2 \le \frac{1}{\sqrt{2}}(y + z)$
$1 \le x_i y_i$	$\left\\|\begin{bmatrix} 2 \\ y_i - x_i \end{bmatrix}\right\\|_2 \le x_i + y_i$
$u^{\frac{3}{2}} \le t$	$vt \ge u^2, u \ge v^2$
$\|x - x_0\|_2 \le \alpha$	$\|[\mathbb{I}][x] - x_0\|_2 \le \alpha$
$\|x - y\|_2 \le \alpha$	$\left\\|[\mathbb{I} \quad -\mathbb{I}] \begin{bmatrix} x \\ y \end{bmatrix}\right\\|_2 \le \alpha$
Robust Linear Programming (LP)	
$\min_x c^T x$	subject to $\forall a_i \in \mathcal{U}, a_i^T x \le b_i$ for $i = 1, \dots, m$
Scenario Uncertainty Model ($\mathcal{U}$ finite set)	
$\mathcal{U}_i = \left\{a_i^{(1)}, \dots, a_i^{(K)}\right\}$ $\forall a_i \in \mathcal{U}_i, a_i^T x \le b_i$	$\min_x c^T x$ subject to $\left(a_i^{(j)}\right)^T x \le b_i$ for $i = 1, \dots, m$
Box Uncertainty Model (LP)	
$\mathcal{U} = \bigcup_{i=1}^{m} \mathcal{U}_i$ $\mathcal{U}_i = \{a : \|a - \hat{a}_i\|_\infty \le \rho\}$	$\max_{a \in \mathcal{U}} a^T x \le b$ $\max_{a \in \mathcal{U}} a^T x = \hat{a}^T x + \max_{\\|y\\|_\infty \le \rho} y^T x = \hat{a}^T x + \rho\|x\|_1$ $\hat{a}^T x + \rho\|x\|_1 \le b$ (LP constraint)
$\min_x c^T x$ subject to $\forall a_i \in \mathcal{U}_i, a_i^T x \le b_i, i = 1, \dots, m$	$\min_x c^T x$ subject to $\hat{a}_i^T x + \rho\|x\|_1 \le b_i, i = 1, \dots, m$
Ellipse Uncertainty Model (SOCP)	
$\mathcal{U} = \{a : (a - a_0)^T P^{-1} (a - a_0) \le 1\}, P \succ 0$ $\mathcal{U} = \{a = \hat{a} + Ru : \|u\|_2 \le 1\}$ where $P = RR^T$, $u = R^{-1}(a - a_0)$	$\max_{a \in \mathcal{U}} a^T x \le b$ $\max_{a \in \mathcal{U}} a^T x = \hat{a}^T x + \max_{u : \\|u\\|_2 \le 1} u^T R^T x$ $\le \hat{a}^T x + \|R^T x\|_2$
$\min_x c^T x$ subject to $\forall a_i \in \mathcal{U}_i, a_i^T x \le b_i, i = 1, \dots, m$	$\min_x c^T x$ subject to $\hat{a}_i^T x + \left\\|R_i^T x\right\\|_2 \le b_i, i = 1, \dots, m$
Robust Optimization	
$\min_x f_0(x)$ subject to $f_i(x) \le 0$ for $i = 1, \dots, m$	$\min_x \max_u F_0(x, u)$ subject to $F_i(x, u) \le 0 \ \forall u \in \mathcal{U}$ for $i = 1, \dots, m$
Robust Least Squares (SOCP)	
$\min_x \max_{\\|\Delta\\|_2 \le \rho} \left\\|(\hat{A} + \Delta)x - y\right\\|_2$	$\Delta = \frac{\rho(\hat{A}x - y)x^T}{\\|Ax - y\\|_2 \\|x\\|_2}$ (rank 1 matrix) $\min_x \left\\|\hat{A}x - y\right\\|_2 + \rho\|x\|_2$ (SOCP)

Vocabulary	
Regression	Measure of relation between one variable and another.
Regularized	Already have a target in mind; want a solution close to the target.
Sparsity	As few nonzero elements as possible; see LASSO.
Robust	A solution that accounts for the uncertainty; ambiguity or error exists in data
Weighted	Just introduce weights; usually no change in problem type
Underdetermined	$m \leq n$, full row rank, AA^T invertible, $N(A^T) = \{0\}$, rows of A linearly independent
Overdetermined	$n \leq m$, full column rank, $A^T A$ invertible, $N(A) = \{0\}$, columns of A linearly independent

Identification	
Least Squares	Slogan: *squared l_2 norm + linear inequality constraints*
LP	l_1 norm/regression, l_∞ norm/regression, box uncertainty, resource management, max flow
QP	Energy, variance, error (squared), index tracking), l_1 regularized least squares (sparsity, piecewise fitting)
SOCP	Ellipsoid uncertainty, inverses, rational power, distance (no square), route planning, sum of norms, general robust optimization

Theory

QR Decomposition $A = QR$

- Standard QR (for full column rank A)

$$A = QR$$

$A \in \mathbb{R}^{m\times n}$ full column rank
$Q \in \mathbb{R}^{m\times n}$ columns orthogonal (i.e. $Q^TQ = \mathbb{I}_n$) and forms an orthogonal basis in $R(A)$
$R \in \mathbb{R}^{n\times n}$ upper triangular, square, invertible.

$A = QR$	$R(A) = R(Q)$
$QQ^Ty = \Pi_{R(A)}(y)$	$A^{pi} = R^{-1}Q^T$

- Full QR (for full column rank A)

$$A = QR = [Q_1 \quad Q_2]\begin{bmatrix} R_1 \\ 0 \end{bmatrix}$$

$A \in \mathbb{R}^{m\times n}$ full column rank
$Q \in \mathbb{R}^{m\times m}$ orthogonal matrix; Q_1 forms orthonormal basis for $R(A)$, Q_2 forms orthonormal basis for $R(A)^\perp = N(A^T)$
$R \in \mathbb{R}^{m\times n}$ with $R_1 = \mathbb{R}^{n\times n}$ upper-triangular and invertible.

In $\|Ax - b\|_2$, $\|Q_2^Tb\|_2$ corresponds to distance between b and orthogonal projection to $R(A)$.

- Not full column rank A

$$AP = QR = Q[R_1 \quad R_2]$$
$$R = [R_1 \quad R_2]P^T$$

$Q \in \mathbb{R}^{m\times r}$ is has orthogonal columns ($Q^TQ = \mathbb{I}_r$)
$R_1 \in \mathbb{R}^{r\times r}$ is upper-triangular, square, invertible
$R_2 \in \mathbb{R}^{r\times(n-r)}$
$P \in \mathbb{R}^{n\times n}$ permutation matrix

- Full QR Decomposition

$$AP = QR = [Q_1 \quad Q_2]\begin{bmatrix} R_1 & R_2 \\ 0 & 0 \end{bmatrix}$$

$Q \in \mathbb{R}^{m\times m}$ orthogonal, square $Q^TQ = QQ^T = \mathbb{I}_m$
$R_1 \in \mathbb{R}^{r\times r}$ upper triangular and invertible
$P \in \mathbb{R}^{n\times n}$ permutation matrix.

Singular Value Decomposition $A \in \mathbb{R}^{m\times n}$

$$A = \sum_{i=1}^{r} \sigma_i u_i v_i^T = U\Sigma V^T = U_r\tilde{\Sigma}V_r^T, \Sigma := \begin{bmatrix} \tilde{\Sigma} & 0 \\ 0 & 0 \end{bmatrix}$$

$$r \le \min(n,m) \;\; \tilde{\Sigma} = diag(\sigma_1, \dots, \sigma_r), \;\; \sigma_1 \ge \dots \ge \sigma_r > 0$$

- $Av_i = \sigma_i u_i$; $A^Tu_i = \sigma_i v_i$; $u_i^TA = \sigma_i v_i^T$
- r is the rank of A, A^T, A^TA and AA^T
- $\sigma_1^2, \dots, \sigma_r^2$ are eigenvalues of AA^T and A^TA
- $u_1, \dots, u_r$ are the eigenvectors for AA^T
- $v_1, \dots, v_r$ are the eigenvectors for A^TA
- $\{u_1, \dots, u_r\}$: an orthonormal basis of $R(A)$

Pseudo-inverse A^{pi} and Square Roots $\sqrt{A}$

Moore-Penrose Pseudo-inverse $A^{pi} \in \mathbb{R}^{n\times m}$

$AA^{pi}A = A$ $A^{pi}AA^{pi} = A^{pi}$ $(AA^{pi})^{pi} = AA^{pi}$ $(A^{pi}A)^{pi} = A^{pi}A$	$\Pi_{N(A)^\perp} = A^{pi}A$ $\Pi_{N(A)} = I - A^{pi}A$ $\Pi_{R(A)} = AA^{pi}$ $\Pi_{R(A)^\perp} = I - AA^{pi}$

Such a matrix always exists and is unique

- A full column rank $A^{pi} = (A^TA)^{-1}A^T$
- A full row rank $A^{pi} = A^T(AA^T)^{-1}$
- A invertible $A^{pi} = A^{-1}$
- A general $A^{pi} = V\Sigma^{pi}U^T = V_r\tilde{\Sigma}^{-1}U_r^T$

$$\Sigma^{pi} = \begin{bmatrix} \sigma_1^{-1} & & \\ & \ddots & \\ & & \sigma_r^{-1} \end{bmatrix}$$

- A positive definite $\Rightarrow \sqrt{A}$ also positive definite

Dual Norms and Matrix Norms

Dual norm $\|\cdot\|^*: \mathbb{R}^n \to \mathbb{R}$:

$$\|y\|^* := \max_{x:\|x\|\le 1} y^Tx \equiv \max_{x:\|x\|\le 1} |y^Tx|$$

Operator norm:

$$\|f\|_{op} := \max_{x:\|x\|\le 1} f(x), \qquad f \in \chi^*$$

Condition number $\kappa(A) = \frac{\sigma_1}{\sigma_n} = \frac{\sigma_{\max}}{\sigma_{\min}}$

If A is square, $\kappa(A) = \|A\|_{LSV}\|A^{-1}\|_{LSV}$
If A is unitary, $\kappa(A) = 1$

Induced p-norms (measure of stretch):

$$\|A\|_p := \max_{x:\|x\|_p\le 1} \|Ax\|_p = \max_{x:\|x\|_p\ne 0} \frac{\|Ax\|_p}{\|x\|_p}$$

- $\|A\|_1 = \max_{1\le i\le n} \|a_i\|_1$ (max l_1 norm of columns)
- $\|A\|_2 = \sqrt{\lambda_{\max}(A^TA)}$
- $\|A\|_\infty = \max_{1\le i\le m} \|a_i^T\|_1$ (max l_1 norm of rows)
- $\|A\|_F = \sqrt{tr(A^TA)} = \sqrt{\sum_{i=1}^n \lambda_i(A^TA)}$

Matrix norms are sub-multiplicative:

$$\|AB\|_p \le \|A\|_p\|B\|_p$$

Empirical Covariance Matrix

$$\Sigma = \mathbb{E}[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^T] = \frac{1}{n}\sum_{i=1}^n (x_i - \hat{x})(x_i - \hat{x})^T$$

$$\Sigma = \frac{1}{n}X^C(X^C)^T = \frac{1}{n}\sum_{i=1}^n x_i x_i^T$$

- $\{u_{r+1}, \dots, u_m\}$: an orthonormal basis of $N(A^T)$
- $\{v_1, \dots, v_r\}$: an orthonormal basis of $R(A^T)$
- $\{v_{r+1}, \dots, v_n\}$: an orthonormal basis of $N(A)$

Full column rank, $r = n$, so $A = U \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} V^T$

Full row rank matrices, $r = m$, $A = U[\Sigma \quad 0]V^T$

$Px = q$ unique solution if Σ full column rank

- Symmetric, positive semi-definite
- Total variance: $tr(\Sigma) = \frac{1}{n}\|X^C\|_F^2$
- kth order explained variance $= \frac{\sigma_1^2 + \dots + \sigma_k^2}{\sigma_1^2 + \dots + \sigma_n^2}$
- Mean along vector u
$$\hat{z} = u^T \hat{x}$$
- Variance along vector u
$$\sigma^2(u) = \frac{1}{m}\sum_{k=1}^{m}\left(u^T(x_k - \hat{x})\right)^2 = u^T \Sigma u$$
- u_1 (the first eigenvector of $X^C(X^C)^T$) maximizes the variance.

Principle Component Analysis

Algorithm

- Center data matrix $X \to X^C$
- Project onto a well-chosen direction $u_1 \in \mathbb{R}^n$ with $\|u\|_2 = 1$ that minimizes the component of data not explained by u_1
$$u_1 = \underset{u_1:\|u_1\|_2 \le 1}{\arg\min} \sum_{i=1}^{m} \|(u_1^T \hat{x}_i)u_1 - \hat{x}_i\|_2^2$$
$$= \underset{u_1:\|u_1\|_2 \le 1}{\arg\max}\ u_1^T \left(\frac{1}{m} X^C (X^C)^T\right) u_1 = v_{\max}(\Sigma)$$
- Subtract the component of X^C explained by u_1 and repeat the algorithm.
- u_1 is the direction of maximum variance, also called the principle component vector

Mechanics $X \in \mathbb{R}^{m \times n}$

- Column average: $\hat{x} = \frac{1}{n} X \mathbb{1}$
- Centering matrix: $P = I - \frac{1}{n}\mathbb{1}\mathbb{1}^T$
$$XP = X^C = X - \hat{x}\mathbb{1}^T$$
$$P\mathbb{1} = 0$$
- For a centered matrix X^C: $X^C \mathbb{1} = 0$
$$\frac{1}{2}\sum_{1 \le i,j \le n} \left\|x_i - x_j\right\|_2^2 = n\,\mathrm{tr}(XPX^T)$$

Properties of Norms and Traces

Norms

- $|x^T y| \le \|x\|_p \|y\|_q, \frac{1}{p} + \frac{1}{q} = 1$
- $\frac{1}{\sqrt{n}}\|x\|_2 \le \|x\|_\infty \le \|x\|_2 \le \|x\|_1 \le \sqrt{n}\|x\|_2 \le n\|x\|_\infty$
- For $x \neq 0$, $\mathrm{card}(x) \ge \frac{\|x\|_1^2}{\|x\|_2^2}$
- $\|A\|_F^2 = \mathrm{tr}(A^T A) = \sum_{i=1}^r \sigma_i^2$
- $\|A + B\|_F^2 = \|A\|_F^2 + \|B\|_F^2 + 2\,\mathrm{tr}(A^T B)$

Traces

- $\mathrm{tr}(A)^2 \le \mathrm{rank}(A)\|A\|_F^2$
- $\mathrm{tr}(AB) \le \|A\|_F \|B\|_F$
- $\mathrm{tr}(u^T u) = \|u\|_F^2 = \|u\|_2^2$
- $\mathrm{tr}(AB) = \mathrm{tr}(BA)$
- $\mathrm{tr}(ba^T) = a^T b$
- $\mathrm{tr}(ABCD) = \mathrm{tr}(BCDA) = \mathrm{tr}(CDAB) = \mathrm{tr}(DABC)$
- $\|AB\|_p \le \|A\|_p \|B\|_p$
- $\|A\|_p \coloneqq \max_{x:\|x\|_p \le 1} \|Ax\|_p = \max_{x:\|x\| \neq 0} \frac{\|Ax\|_p}{\|x\|_p}$
- $\|x\|_2^2 = \|Ux\|_2^2$ for U orthogonal
- $x^T Q x = tr(x^T Q x) = tr(x x^T Q)$

Polyhedral

Definition: A function is <u>polyhedral</u> if its epigraph is polyhedral i.e. if and only if $\exists C \in \mathbb{R}^{m \times (n+1)}, d \in \mathbb{R}^m$ s.t.
$$\mathrm{epi}\, f = \{(x,t) \in \mathbb{R}^{n+1} : t \ge f(x)\}$$
$$= \left\{(x,t) \in \mathbb{R}^{n+1} : C\begin{bmatrix} x \\ t \end{bmatrix} \le d\right\}$$

[Max affine functions] Functions expressible as the max of a finite number of affine functions is polyhedral $f(x) = \max_{1 \le i \le m} a_i^T x + b_i$

<u>Consequence</u>: l_∞ norm is polyhedral.

Functions expressible as the sum of functions that are max affine functions are polyhedral.
<u>Consequence</u>: l_1 norm is polyhedral.
The projection of polyhedral is polyhedral.

Problem Solving Techniques

- Take derivatives!
- Fix a few variables, minimize w.r.t. the rest.
- Reduce to classic problems/formulations
- Group terms and simplify
- Geometry argument / Decomposition
- Enjoy long problems (your strength!)

Final Checks

- Sign constraints? $x \ge 0$
- Standard form
- Check symmetric property for positive semi-definite

Appendix I: Mathematical Toolbox

Linear algebra, multivariable calculus tricks and techniques

Linear Algebra

$(\mathrm{rank}(A))^\perp = \mathrm{null}(A^T)$	$(\mathrm{null}(A))^\perp = \mathrm{rank}(A^T)$
$(\mathrm{rank}(A^T))^\perp = \mathrm{null}(A)$	$(\mathrm{null}(A^T))^\perp = \mathrm{rank}(A)$

$$S \subset T \Rightarrow T^\perp \subset S^\perp$$

Projection

- Let $\{u_1, \dots, u_n\}$ be an orthonormal basis for V. Then $P = \Sigma u_i u_i^T$ is the matrix of orthogonal projection.
- If A is a matrix with its column as any basis of V, then $P = A(A^T A)^{-1} A^T$ is the matrix of orthogonal projection.

Decompositions

- [Spectral] Any symmetric matrix has exactly n **real** (not necessarily distinct) eigenvalues; and eigenvectors can be chosen to be orthonormal.
$$A = U\Lambda U^T = \Sigma_i \lambda_i u_i u_i^T$$
- [Cholesky] If A is symmetric positive definite, then $A = LL^T$ where L is lower triangular with real and positive diagonal entries (i.e. L invertible)
- A symmetric positive definite $\Rightarrow$ every eigenvalue is positive (similar for semi)
- A positive definite, then $B^T AB$ also positive definite (similar for semi)
- A positive definite, then exists unique positive definite $\sqrt{A}$. (similar for semi)
$$\sqrt{A} = U\sqrt{\Lambda}U^T$$
- [Inertia Theorem] A symmetric matrix is congruent to a diagonal matrix with $0, 1, -1$ under $A = B\Lambda B^T$.
- [Sylvester's Rule of Inertia] Negative index of inertia $q =$ the number of sign changes of the leading minors $\Delta_0 = 1, \dots, \Delta_n = \det A$
- [Schur Complement] Let S be a symmetric matrix partitioned into blocks $S = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix}$ with C positive definite. TFAE:
    - S is positive semi-definite.
    - $A - BC^{-1}B^T$ (Schur's complement of C) is positive semi-definite

Multivariable Calculus

Determinants

$$\det(\mathbb{I} + uv^T) = 1 + u^T v$$
$$\det(\mathbb{I}_2 + A) = 1 + \det A + tr(A)$$
$$\det(I + \epsilon A) \approx 1 + \det(A) + \epsilon\, tr(A) + \frac{1}{2}\epsilon^2 tr(A)^2 - \frac{1}{2}\epsilon\, tr(A^2)$$

Approximations

$$f(x) \approx f(x_0) + \nabla f(x_0)^T (x - x_0)$$
$$f(x) \approx q(x) = f(x_0) + \nabla f(x_0)^T (x - x_0) + \frac{1}{2}(x - x_0)^T \nabla^2 f(x_0)(x - x_0)$$

Derivatives

$$\nabla(a^T x + b) = a \qquad \nabla(x^T x) = 2x$$
$$\nabla_X \|X - P\|_F^2 = 2(X - P)$$

$$\nabla^2 \left(\frac{1}{2}x^T Ax\right) = \frac{1}{2}(A + A^T)$$
$$g(x) = f(Ax + b) \Rightarrow \nabla g(x) = A^T \nabla f(Ax + b)$$

$$q(x) = \frac{1}{2}\begin{bmatrix} x & 1 \end{bmatrix} \begin{bmatrix} A & b \\ b^T & c \end{bmatrix} \begin{bmatrix} x \\ 1 \end{bmatrix} = \frac{1}{2}x^T Ax + b^T x + c$$
$$\nabla q = Ax + b$$
$$\nabla^2 q = A$$

$$\nabla_X \mathrm{tr}(XHX^T) = XH^T + XH$$
$$\nabla_X \mathrm{tr}(AX) = A^T$$

$$H_q = \begin{bmatrix} \dfrac{\partial^2 q}{\partial x_1^2} & \dfrac{\partial^2 q}{\partial x_1 \partial x_2} \\ \dfrac{\partial^2 q}{\partial x_2 \partial x_1} & \dfrac{\partial^2 q}{\partial x_2^2} \end{bmatrix}$$

Inequalities

- Cauchy Schwarz / Holder's $p^{-1} + q^{-1} = 1$:
$$|x^T y| \le \|x\|_p \|y\|_q \Rightarrow \max_{y:\|y\|_p \le 1} y^T x = \|x\|_q$$

Equality when $y_i = \frac{sign(x_i)|x_i|^{p-1}}{\|x\|_q}$

- Power Mean Inequality
$$p \ge q \Rightarrow n^{-\frac{1}{p}}\|x\|_p \ge n^{-\frac{1}{q}}\|x\|_q$$
- Smoothing
- Sherman-Morrison-Woodbury
$A \in \mathbb{R}^{n\times n}, u, v \in \mathbb{R}^n$ s.t. $A, A + uv^T$ non-singular
$$(A + uv^T)^{-1} = A^{-1} - \frac{A^{-1}uv^T A^{-1}}{1 + v^T A^{-1} u}$$
- Minkowski's Inequality $1 \le p < \infty$
$$\|f + g\|_p \le \|f\|_p + \|g\|_p$$

Appendix II: Diagrams