

NLA: Golub for $k = 1 : m, n$: $u_k = (\text{sgn}(b_{k,k})\|b_{k:m,k}\|e_1 + b_{k:m,k})$; $u_k := \hat{u}_k$; $U_k := I - 2u_k u_k^T$;
 $B_{k:m,k:n} := U_k B_{k:m,k:n}$; $U = [I_{k-1,k-1}, 0; 0, U_k]$; **for** $j = 1 : m, n - 1$: $v_k^T := \text{sgn}(b_{k,k+1})\|b_{k,k+1:n}\|e_1 +$
 $b_{k:m,k}; V_k := I - 2v_k v_k^T$; $B_{1:m,k+1:n} = B_{1:m,k+1:n} V_k$; $V = [I_{k,k}, 0; 0, V_k]$ **endfor endfor**; $2 \cdot (2mn^2 - 2n^3/3)$
Householder for $k = [1, n]$: $x = A_{k:m,k}; v_k = \text{sgn}(x)\|x\|e_k + x$; $v_k = \frac{v_k}{\|v_k\|}$ **for** $j = [k, n]$ $A_{k:m,j} =$
 $A_{k:m,j} - 2v_k [v_k^* A_{k:m,j}]$ **endfor endfor**. $2mn^2 - \frac{2n^3}{3}$. **MG-S** $V = A$; **for** $i = [1, n]$: $r_{ii} = \|v_i\|$; $q_i = \frac{v_i}{r_{ii}}$; **for**
 $j = [i + 1, n]$ $v_j = v_j - (q_i^T v_j) q_i$; $r_{ij} = q_i^T v_j$ **endfor endfor**. $2mn^2$. **Arnoldi**: $q_1 := \hat{b}$; $q_{k+1} h_{k+1,k} =$
 $Aq_k - \sum_{i=1}^k q_i h_{ik}$; $h_{ik} = q_i^T (Aq_k)$; $h_{k+1,k} := \|v\| \rightarrow Aq_k := Q_k H_k + q_{k+1} [0 \dots h_{k+1,k}]$. **Givens**
 $3mn^2$ **SVD**: $= \sum_i^{\min m, n} u_i \sigma_i v_i^T$. **C-F**: $\sigma_i(A) \sigma_n(B) \leq \sigma_i(AB) \leq \sigma_i(A) \sigma_1(B)$ **QR Algo**: $A_{k+1} =$
 $Q_k^T A_k Q_k \rightarrow A_{k+1} = (Q^{(k)})^T A Q^{(k)}$. Next $A^{k-1} = (Q_1 \dots Q_{k-1})(R_{k-1} \dots R_1)$, so $A_k = Q_k R_k =$
 $(Q^{(k-1)})^T A Q^{(k-1)}$ so $Q^{(k-1)} A_k = A Q^{(k-1)}$. So $A^k = (A Q^{(k-1)}) R^{(k-1)} = Q^{(k)} R^{(k)}$ as $A_k = Q_k R_k$.
Krylov: Usually want $x_k - x_0 \in \mathcal{K}_k$. Also note to show **CG span properties** first show $\{r_k\} \subset \{p_k\}$ etc,
then show $\{p_k\}$ LI. **GMRES**: $\min \|AQ_k y - b\|_2 \rightarrow \min \|H_k y - \|b\|e_1\|$. Bound $\|r_k\| = \|Mp(\Lambda)M^{-1}r_0\|$
GMRES Conv: If $x_k = p_{k-1}(A)b$ have $\min \|Ax_k - b\| = \min_{p(0)=1} \|Ap_{k-1}(A)b - b\| \leq k_2(A)\|p(\Lambda)b\|$
with $p(0) = 1$ **CG Bound**: With $c = x - x_0, c_k = x_k - x_0$ s.t. $r_k = A(c - c_k)$ we have $r_k^T v =$
 $0 \forall v \in \mathcal{K}_k$ so $v^T A(c - c_k) = 0$, s.t. $y = c_k = \arg \min \|c - y\|_A$. WTS $e_k = e_0 p_k(A)$ with $p(0) = 1$,
and write $e_0 := \sum \gamma_i v_i$ with $Av_i = \lambda_i v_i \rightarrow \|e_k\|_A = \min_{p_k, p(0)=1} \max |p(\lambda_i)| \|e_0\|_A$ **CG Conver-**
gence: $\|e_k\|_A = \min_{p(0)=1} \|p_k(A)e_0\| = \min_{p_k(A)} \max |p_k(\lambda)| \|e_0\| \rightarrow \leq 2((\sqrt{k_2} - 1)/(\sqrt{k_2} + 1))^k$; need
 $\alpha := 2(\lambda_1 + \lambda_2)$ **Cheb**: $T_k(x) = \frac{1}{2}(z^k + z^{-k})$; $2xT_k = T_{k+1} + T_{k-1}$ **Cheb Shift**: Choose $p(x) =$
 $T_k([2x - b - a]/[b - a])/T_k([-b - a]/[b - a])$ s.t. $p(0) = 1$. Then $p \leq 1/|T_k([-b - a]/[b - a])| \leq$
 $2([\sqrt{\kappa} - 1]/[\sqrt{\kappa} + 1])^k$ **CG Conditions**: To show $r_{k+1}^T r_k = 0$ first show $p_k^T A p_k = p_k^T A r_k$ via β then show
 $p_k^T r_k = r_k^T r_k$ via $p_{k-1}^T r_k = 0$. **Preconditioning**: For GMRES solve $MAx = Mb$ with MA eigvals clus-
tered far from 0, well conditioned. E.g. if $A = LU$ do $(LU)^{-1}$. For CG, $A = A^T$ difficult so want $M^T M \approx$
 A^{-1} . So want $M^T A M y = M^T b$, same properties. **MP**: $\sigma(G) \in [\sqrt{m} - \sqrt{n}, \sqrt{m} + \sqrt{n}] \rightarrow k_2 = O(1)$
Sketch: with $GA\hat{x} = Gb$, and via $C - F \|G[A, b][v, -1]^T\| \leq (s + \sqrt{n+1})\|R[v, -1]^T\|$, similar for lower
bound via MP $\rightarrow \|A\hat{x} - b\| \leq (\sqrt{s} + \sqrt{n+1})/(\sqrt{s} - \sqrt{n+1})\|Ax - b\|$ **Blend**: solve $\|(A\tilde{R}^{-1})y - b\| = 0$
via CG; $k_2(A\tilde{R}^{-1}) = O(1)$ with $GA = \tilde{Q}\tilde{R}$ **PROOF**: $A = QR$; $GA = GQR = \hat{G}R$. Let $\hat{G} = \hat{Q}\hat{R}$ so
 $GA = \hat{Q}\hat{R}R = \hat{Q}(\hat{R}R) \rightarrow \hat{R}^{-1} = R^{-1}\hat{R}^{-1} \rightarrow k_2(A\hat{R}^{-1}) = k_2(\hat{R}^{-1}) = O(1)$ by MP. $O(mn)$ to solve via
normal **HMT**: For $X = n \times r$ let $AX = QR$, then if $A = U_r \Sigma_r V_r^T$, $\text{span}(Q) = \text{span}(U_r)$ so $\hat{A} = QQ^T A$ is a
rank r approximant. **HMT Proof**: Goal $\|A - \hat{A}\| = O(1)\|A - A_r\|$. Have $(I - QQ^T)AX = 0$ so $A - \hat{A} =$
 $(A_n - QQ^T)A(I_n - XM^T) = 0 \forall M^T$. Choose $M^T = (V^T X)^\dagger V^T$, $V \in n \times \hat{r} \leq r$. Let $XM^T = P$ s.t. $A(I -$
 $P) = A(I - VV^T)(I - P)$. So $\|A - \hat{A}\| = \|(I_m - QQ^T)U_A \Sigma_A [\tilde{V}_r^T, \tilde{V}_{r+1}^T]^T (I - VV^T)(I - P)\|$. If $V =$
 \tilde{V}_r then $= \|(I_m - QQ^T)U_A \Sigma_A [0, \tilde{V}_{r+1}^T]^T (I - P)\| \leq \|\Sigma_{\hat{r}+1}\| \|I_n - XM^T\|$. Now note $\|I_n - XM^T\| =$
 $\|I_n - X(\tilde{V}_r^T X)^\dagger \tilde{V}_r^T\| \leq 1 + \|X\| \|(\tilde{V}_r^T X)^\dagger\|$. Now $\|X\| \leq \sqrt{n} + \sqrt{r}$ by MP, and $\|(\tilde{V}_r^T X)^\dagger\| = \sigma_n(\tilde{V}_r^T X)^{-1} \leq$
 $(\sqrt{r} - \sqrt{\hat{r}})^{-1}$ by MP. So $\|XM^T\| \leq \frac{\sqrt{n} + \sqrt{r}}{\sqrt{r} - \sqrt{\hat{r}}}$. **Bounds**: $\|ABB^{-1}\| \geq \|AB\| \|B^{-1}\| \rightarrow \|A\|/\|B^{-1}\| \geq$
 $\|AB\|$. **Weyls**: $\sigma_i(A + B) = \sigma_i(A) + [-\|B\|, \|B\|]$ **Rev Δ Ineq**: $\|A - B\| \geq \|A\| - \|B\|$ **Courant**
Application: $\sigma_i([A_1; A_2]) \geq \max(\sigma_i(A_1), \sigma_i(A_2))$ **Schur**: Take $Av_1 = \lambda_1 v_1$; construct $U_1 = [v_1, V_\perp] \rightarrow$
 $AU_1 = U_1[e_1, X]$. Repeat. **Conditioning** $\kappa_2(A) = \sigma_1/\sigma_n = \|A\| \|A^{-1}\|$ **Similarity**: $A \rightarrow P^{-1}AP$, same
 λ . **Pseud-Inv**: $A^\dagger = V \Sigma^{-1} U^T$
CO: G-N: $\vec{x}_{k+1} = \vec{x}_k - \frac{\nabla f_k}{J^T J}$, with $J :=$ Jacobian of $r(x)$ **Linesearch Convergence**: Show $x_{k+1} - x_* =$
 $\Psi(x_k) - x_* = \Psi(x_* + e_k) - x_*$ and Taylor expand. **SD**: $\|x_{k+1} - x_*\| \leq (k_2(H) - 1)/(k_2(H) + 1)\|x_k - x_*\|$
with H hessian. Also note with EXACT linesearch for quadratic, $H(x - x_*) = -s$. **Rayleigh**:
 $\frac{s^T H s}{\|s\|^2} \leq \|H\|$ **bArm**: To show existence of α , have $\phi(\alpha) = f(x_k + \alpha s_k)$, $\psi(\alpha) = \phi(\alpha) - \phi(0) - \beta \alpha \phi'(0) \leq 0$,
show $\psi'(0) = (1 - \beta)\phi'(0) \leq 0 \rightarrow \psi(\alpha) \downarrow$ with α . **BFGS**: To show $H_{k+1} \geq 0$ nec. $\gamma^T \delta > 0$. Suff via γ, δ
LI \rightarrow use $\|\cdot\|_H \rightarrow \gamma^T \delta > 0$. **Quad Penalty Meth** With $y = -c/\sigma$, $\|\nabla_\sigma \Phi\| \leq \epsilon^k, \sigma^k \rightarrow 0, x \rightarrow x_*$, $\nabla c(x_*)$
LI, then $y \rightarrow y_*$, $x \rightarrow KKT$, if $f, c \in C^1$. **PROOF**: If $y_* := J_*^\dagger \nabla f_* \rightarrow \|y_k - y_*\| = \|J_k^\dagger \nabla f_* - y_k\| \leq$
 $\|J_k^\dagger \nabla f_k - J_k^\dagger \nabla f_*\| + \|J_k^\dagger \nabla f_k - y_k\|$. Next $\|J_k^\dagger \nabla f_k - y_k\| \leq \|J_k^\dagger\| \|\nabla_\sigma \Phi\| \rightarrow 0$. Also, $\nabla f_* - J_*^T y_* \rightarrow 0$,
and $c_{k \rightarrow *} = -\sigma^{k \rightarrow *} y_{k \rightarrow *} = 0$ so $x_* \rightarrow KKT$ **Quad Pen. Meth Newt** Have $w = (J \Delta x + c)/\sigma$ so
 $[\nabla^2 f, J^T; J, -\sigma I][\Delta x, w]^T = -[\nabla f, c]$ **Trust Region Radius**: $\rho_k := (f(x_k) - f(x_k + s_k))/(f(x_k) -$
 $m_k(s_k))$ **TR-Method**: If $\rho \geq 0.9$ then double radius, update step $x_{k+1} = x_k + s_k$. If $\rho \geq 0.1$ then
same radius, update step. If ρ small shrink radius, don't update step. **Cauchy**: Is the point on gradient
which minimises the quadratic model within TR. Want $m_k(s_k) \leq m_k(s_{kc})$, where $s_{kc} := -\alpha_{kc} \nabla f(x_k)$,

and $\alpha_{kc} := \arg \min m_k(\alpha \nabla f(x_k))$ subject to $\|\alpha \nabla f\| \leq \Delta$, i.e. $\alpha_{max} := \Delta / \|\nabla f\|$. **Calculation of Cauchy:** We want to prove cauchy model decrease i.e. $f(x_k) - m_k(s_k) \geq f(x_k) - m_k(s_{kc}) \geq 0.5 \|\nabla f_k\| \min \left\{ \Delta_k, \frac{\|\nabla f_k\|}{\|\nabla^2 f_k\|} \right\}$. First define $\Psi(\alpha) := m_k(-\alpha \nabla f)$ s.t. $\Psi := f_k - \alpha \|f_k\|^2 - 0.5 \alpha^2 H_k$, with $H_k := [\nabla f_k]^T [\nabla^2 f_k] [\nabla f_k]$. N.B. that $\alpha_{min} := \frac{\|\nabla f_k\|^2}{H_k}$ if $H_k > 0$, from $\Psi'(0) < 0$. Now **A: If $H_k \leq 0$** then we have $\Psi(\alpha) \leq f_k - \alpha \|\nabla f_k\|^2 \rightarrow \alpha_{kc} = \alpha_{max}$. So, we have $f_k - m_{s_k} \geq f_k - m_{s_{kc}} \geq \|\nabla f_k\| \Delta_k \geq 0.5 \|\nabla f_k\| \min \{\Delta_k\}$. Now **B.i: If $H_k > 0 \rightarrow \alpha_{kc} = \alpha_{min}$** . Here $f_k - m_{s_{kc}} = \alpha_{kc} \|\nabla f\|^2 - 0.5 \alpha_{kc}^2 H_k = \frac{\|\nabla f\|^4}{2H_k} \geq \frac{\|\nabla f\|}{2} \min \left\{ \frac{\|\nabla f\|}{\|\nabla^2 f\|} \right\}$ via C-S. Now **B.ii: If $H_k > 0 \rightarrow \alpha_{kc} = \alpha_{max}$** . Here $\Delta / \|\nabla f\| \leq \|\nabla f\|^2 / H_k \rightarrow \alpha_{kc} H_k \leq \|\nabla f\|^2$. So $f_k - m_{kc} = -\alpha_{kc} \|\nabla f\|^2 + \frac{\alpha_{kc}^2}{2} H_k \geq \frac{\|\nabla f\|^2}{2} \alpha_{kc} \geq 0.5 \|\nabla f\| \min \{\Delta_k\}$ **TR-Global Convergence:** If $m_k(s_k) \leq m_k(s_{kc})$ then either $\exists k \geq 0$ s.t. $\nabla f_k = 0$ or $\lim \|\nabla f\| \rightarrow 0$. Further, require $f \in C^2$, bounded below and also ∇f L-cont. **PROOF:** Using def of ρ , $f_k - f_{k+1} \geq \frac{0.1}{2} \|\nabla f_k\| \min \{\dots\}$ from above. Let $\|\nabla^2 f\| := L$, and assuming $\|\nabla f\| \geq \epsilon$ we have $f_k - f_{k+1} \geq 0.05 \frac{\epsilon}{L} \epsilon^2$ assuming TR has a lower bound $c\epsilon/L$. Then sum over all successful jumps s.t. $f_0 - f_{lower} \geq \sum_{i \in \mathbb{S}} f_i - f_{i+1} \geq |\mathbb{S}| \frac{0.05 c \epsilon^2}{L}$ **Solving TR Prob:** Solve secular $\|s\|^{-1} - \Delta^{-1} = 0$. **KKT Feasibility:** Need $s^T J \geq 0$, $J_E^T s = 0$, and $s^T \nabla f < 0$. **KKT Conditions: REMEMBER $c \geq 0$, $\lambda \geq 0$!** **First Order KKT (Equality):** If we have x_* local min, then let $x = x_* + \alpha s$. Then we have $c_i(x(\alpha)) \rightarrow 0 = c_i(x_*) + \alpha s^T J \rightarrow s^T J = 0$. Further, we have $f(x) = f(x_*) + \alpha s^T \nabla f \rightarrow \alpha s^T \nabla f \geq 0$. Repeat for negative α s.t. $s^T \nabla f = 0$. By Rank-Nullity (assuming $J_E(x_*)$ full rank), we have $\nabla f_* = J_*^T y + s_*$ for some y_* , which then implies (after s^T from LHS) that $\|s_*\| = 0$, so $\nabla f_* = J_*^T y_*$. **KKT 2nd Order** If we have $\min f$ with $c(x) \geq 0$, 2^{nd} order conditions are that $s^T \nabla^2 \mathcal{L} s \geq 0$ for all $s \in \mathcal{A}$, with \mathcal{A} defined s.t.: **EITHER** $s^T J_i = 0 \forall i$ s.t. $\lambda_i > 0, c_i = 0$, **OR** $s^T J_i \geq 0 \forall i$ s.t. $\lambda_i = 0, c_i = 0$, for J, c, λ evaluated at x_* . For EQUALITY constraints instead need positive definite $\forall s$ s.t. $J^T s = 0$ **Convex Problems** $\hat{x} = KKT \Rightarrow \hat{x} = \arg \min f(x)$. Proof via $f \geq f(\hat{x}) + \nabla f^T(x - \hat{x})$ so $f \geq f(\hat{x}) + \hat{y}^T A(x - \hat{x}) + \sum_{i \in I} \lambda_i J_i^T(x - \hat{x})$. Choose $Ax = b$, and note that c_i concave s.t. $\lambda_i J_i^T(\hat{x})(x - \hat{x}) \geq \lambda_i(c_i(x) - c_i(\hat{x})) = \lambda_i c_i(x) \geq 0 \rightarrow f(x) \geq f(\hat{x})$. **Log-Barrier Global Convergence:** (for $f - \sum \mu \log(c_i)$) With $f \in C^1$, $\lambda_{ik} = \frac{\mu_k}{c_{ik}}$, $\|\nabla f_u(x_k)\| \leq \epsilon_k, \mu_k \rightarrow 0, x_k \rightarrow x_*$. Also, $\nabla c(x_*)^T J_i \forall i \in \mathcal{A}$ (active constraints). Then x_* KKT and $\lambda \rightarrow \lambda_*$. **PROOF:** Have $J_A^\dagger(x_*) = (J_A(x_*) J_A(x_*)^T)^{-1} J_A(x_*)$. Also, $c_A = 0, c_I > 0$. So $\lambda = \mu/c \rightarrow 0$ so $\lambda_I = 0$ as $c_I > 0$. Next $\|\nabla f_k - J_{Ak} \lambda_{Ak}\| \leq \|\nabla f_k - J_k^T \lambda_k\| + \|\lambda_I\| M_1 = \|\nabla f_{\mu k}\| \rightarrow 0$. Now $\left\| J_A^\dagger \nabla f_k - \lambda_{kA} \right\| \leq \left\| J_A^\dagger \right\| \left\| \nabla f_k - J_{Ak}^T \lambda_{Ak} \right\| \rightarrow 0$. So with triangle ineq $\left\| \lambda_{kA} - J_{Ak}^\dagger \nabla f_k + J_{Ak}^\dagger \nabla f_k - \lambda_{A*} \right\| \rightarrow 0$, via cont. of ∇f and J^\dagger . Thus $\lambda_{Ak} \rightarrow \lambda_{A*} \geq 0$. Combine s.t. $\nabla f_k - J_{Ak}^T \lambda_{Ak}$ with $k \rightarrow *$ so get KKT. **Primal-Dual Newton:** Have $\nabla f = J^T \lambda, C(x) \lambda = \mu e$ so $[\nabla^2 \mathcal{L}, -J^T; \Lambda J, C][dx, d\lambda]^T = -[\nabla f - J^T \lambda, C \lambda - \mu e]^T$ **Augmented Lagrangian:** Same result as QUAD PEN METH but $x \rightarrow x_*$ if $\sigma \rightarrow 0$ for bounded u_k or u_k to y_* for bounded σ_k . Proof via $\|c_k\| \leq \sigma_k \|y_k - y_*\| + \sigma_k \|u_k - y_*\|$. If u_k bounded then $\rightarrow 0$ as $\sigma \rightarrow 0$, else trivially if $u_k \rightarrow y_*$ then to 0. **GLM Global Convergence:** With $f \in C^1, \nabla f$ L-Cont, and f bdd below, then $\nabla f_l = 0$ or $\lim \min \left\{ \frac{|\nabla f^T s|}{\|s\|}, |s^T \nabla f| \right\} \rightarrow 0$. In non-trivial case, via bArmijo, $f_k - f_{k+1} \geq -\beta \alpha_k s_k^T \nabla f_k = \beta \alpha_k |s_k^T \nabla f_k|$. Sum s.t. $f_0 - f_{k+1} \geq \beta \sum \alpha_k |s_k^T \nabla f_k|$, so term in sum to 0. For all k successful we then have $\alpha_k |s_k^T \nabla f_k| \geq \frac{(1-\beta)\tau}{L} \left(\frac{|s_k^T \nabla f_k|}{\|s_k\|} \right)^2 \geq 0$ so squared term to 0. For unsuccessful steps $\alpha_k \geq \alpha_0$ so no norm term. **Convergence Newton LSearch:** Need $f \in C^2$ then if H_k (hessian) bdd above and below, so $\lambda_n \leq \lambda(H_k) \leq \lambda_1$. So $|s_k^T \nabla f_k| \geq \lambda_1^{-1} \|\nabla f_k\|^2$. Also $\|s_k\|^2 \leq \lambda_n^{-2} \|\nabla f_k\|^2$. Thus $\lim \min \left\{ \lambda_n \lambda_1^{-1} \|\nabla f\|, \lambda_1^{-1} \|\nabla f\|^2 \right\} \rightarrow 0$ from GLM Convergence Thm.