NLA: **Golub** for $k=1:m,n$: $u_k=(sgn(b_{k,k})\|b_{k:m,k}\|e_1+b_{k:m,k})$; $u_k:=\hat{u}_k$; $U_k:=I-2u_ku_k^T$; $B_{k:m,k:n}:=U_kB_{k:m,k:n}$; $U=[I_{k-1,k-1},0;0,U_k]$;for $j=1:m,n-1$: $v_k^T:=sgn(b_{k,k+1})\|b_{k,k+1:n}\|e_1+b_{k:m,k}$;$V_k:=I-2v_kv_k^T$; $B_{1:m,k+1:n}=B_{1:m,k+1:n}V_k$; $V=[I_{k,k},0;0,V_k]$endfor endfor; $2\cdot(2mn^2-2n^3/3)$ **Householder** for $k=[1,n]$: $x=A_{k:m,k}$; $v_k=sgn(x)\|x\|e_k+x$; $v_k=\frac{v_k}{\|v_k\|}$ for $j=[k,n]$ $A_{k:m,j}=A_{k:m,j}-2v_k\left[v_k^*A_{k:m,j}\right]$ 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 $j=[i+1,n]$ $v_j=v_j-(q_i^Tv_j)q_i$; $r_{ij}=q_i^Tv_j$ endfor endfor. $2mn^2$. **Arnoldi:** $q_1:=\hat{b}$; $q_{k+1}h_{k+1,k}=Aq_k-\sum_{i=1}^{k}q_ih_{ik}$; $h_{ik}=q_i^T(Aq_k)$; $h_{k+1,k}:=\|v\|\to AQ_k:=Q_kH_k+q_{k+1}[0\dots h_{k+1,k}]$. **Lanczos:** If $A=A^T$ for arnoldi, $AQ_k=Q_kT_k+q_{k+1}[0\dots t_{k+1,k}]$. T tridiag via $H_k=Q_k^TAQ_k$. Also have $t_{k+1}q_{k+1}=(A-t_{k,k})q_k-t_{k-1,k}q_{k-1}$. **Givens** $3mn^2$ **SVD:** $=\sum_i^{r:=\min m,n}u_i\sigma_iv_i^T$. **C-F:** $\sigma_i(A)\sigma_n(B)\le\sigma_i(AB)\le\sigma_i(A)\sigma_1(B)$ **QR Algo:** $A_{k+1}=Q_k^TA_kQ_k\to A_{k+1}=\left(Q^{(k)}\right)^TAQ^{(k)}$. Next $A^{k-1}=(Q_1\dots Q_{k-1})(R_{k-1}\dots R_1)$, so $A_k=Q_kR_k=(Q^{(k-1)})^TAQ^{(k-1)}$ so $Q^{(k-1)}A_k=AQ^{(k-1)}$. So $A^k=(AQ^{(k-1)})R^{(k-1)}=Q^{(k)}R^{(k)}$ as $A_k=Q_kR_k$. **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:** $\min\|AQ_ky-b\|_2\to\min\|H_ky-\|b\|e_1\|$. Bound $\|r_k\|=\left\|Mp(\Lambda)M^{-1}r_0\right\|$ **GMRES Conv:** If $x_k=p_{k-1}(A)b$ have $\min\|Ax_k-b\|=\min_{p(0)=1}\|Ap_{k-1}(A)b-b\|\le k_2(A)\|p(\Lambda)b\|$ 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 $r_k^Tv=0\ \forall\ v\in\mathcal{K}_k$ so $v^tA(c-c_k)=0$, s.t. $y=c_k=\arg\min\|c-y\|_A$. WTS $e_k=e_0p_k(A)$ with $p(0)=1$, and write $e_0:=\sum\gamma_iv_i$ with $Av_i=\lambda_iv_i\to\|e_k\|_A=\min_{p_k,p(0)=1}\max|p(\lambda_i|\|e_0\|_A$ **CG Convergence:** $\|e_k\|_A=\min_{p(0)=1}\|p_k(A)e_0\|=\min_{p_k(A)}\max|p_k(\lambda)|\|e_0\|\to\le 2\left((\sqrt{k_2}-1)/(\sqrt{k_2}+1)\right)^k$; 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 $p\le 1/|T_k([-b-a]/[b-a])|\le 2([\sqrt{\kappa}-1]/[\sqrt{\kappa}+1])^k$ **CG Conditions:** To show $r_{k+1}^Tr_k=0$ first show $p_k^TAr_k=p_k^TAp_k$ via sub for $r_k=p_k-\beta\dots$ then show $p_k^Tr_k=r_k^Tr_k$ via $p_{k-1}^Tr_k=0$. **Preconditioning:** For GMRES solve $MAx=Mb$ with MA eigvals clustered far from 0, well conditioned. E.g. if $A=LU$ do $(LU)^{-1}$. For CG, $A=A^T$ difficult so want $M^TM\approx A^{-1}$. So want $M^TAMy=M^Tb$, same properties. **MP:** $\sigma(G)\in[\sqrt{m}-\sqrt{n},\sqrt{m}+\sqrt{n}]\to k_2=O(1)$ **Sketch:** First $G=\mathbb{R}^{s\times m}$ so with $\hat{x}=\arg\min\|G(Ax-b)\|$, and via $C-F$ $\left\|G[A,b][v,-1]^T\right\|\le(\sqrt{s}+\sqrt{n+1})\left\|R[v,-1]^T\right\|$. Similarly $\left\|G[A,b][v,-1]^T\right\|\ge((\sqrt{s}-\sqrt{n+1}))^{-1}\left\|R[v,-1]^T\right\|$. Then $\|G(A\hat{x}-b)\|\le\|G(Ax-b)\|$ so $\to\|A\hat{x}-b\|\le(\sqrt{s}+\sqrt{n+1})/(\sqrt{s}-\sqrt{n+1})\|Ax-b\|$. **Blend:** solve $\left\|(A\tilde{R}^{-1})y-b\right\|=0$ 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 $GA=\hat{Q}\hat{R}R=\hat{Q}(\hat{R}R)\to\tilde{R}^{-1}=R^{-1}\hat{R}^{-1}\to k_2(A\tilde{R}^{-1})=k_2(\hat{R}^{-1})=O(1)$ by MP. $O(mn)$ to solve via normal **HMT:** For $X=n\times r$ let $AX=QR$, then if $A=U_r\Sigma_rV_r^T$, span(Q)=span(U_r) so $\hat{A}=QQ^TA$ is a rank r approximant. **HMT Proof:** Goal $\left\|A-\hat{A}\right\|=O(1)\|A-A_r\|$. Have $(I-QQ^T)AX=0$ so $A-\hat{A}=(I_m-QQ^T)A(I_n-XM^T)=0\ \forall\ M^T$. Choose $M^T=(V^TX)^\dagger V^T$, $V\in n\times\hat{r}\le r$. Let $XM^T=P$ s.t. $A(I-P)=A(I-VV^T)(I-P)$. So $\left\|A-\hat{A}\right\|=\left\|(I_m-QQ^T)U_A\Sigma_A[\tilde{V}_{\hat{r}}^T,\tilde{V}_{\hat{r}+1}^T]^T(I-VV^T)(I-P)\right\|$. If $V=\tilde{V}_r$ then $=\left\|(I_m-QQ^T)U_A\Sigma_A[0,\tilde{V}_{\hat{r}+1}]^T(I-P)\right\|\le\|\Sigma_{\hat{r}+1}\|\left\|I_n-XM^T\right\|$. Now note $\left\|I_n-XM^T\right\|=\left\|X(\tilde{V}_r^TX)^\dagger\tilde{V}_r^T\right\|\le\|X\|\left\|(\tilde{V}_r^TX)^\dagger\right\|$. Now $\|X\|\le\sqrt{n}+\sqrt{r}$ by MP, and $\left\|(\tilde{V}_r^TX)^\dagger\right\|=\sigma_n(\tilde{V}_r^TX)^{-1}\le(\sqrt{r}-\sqrt{\hat{r}})^{-1}$ by MP. So $\left\|XM^T\right\|\le\frac{\sqrt{n}+\sqrt{r}}{\sqrt{r}-\sqrt{\hat{r}}}$. **Bounds:** $\left\|ABB^{-1}\right\|\ge\|AB\|\left\|B^{-1}\right\|\to\|A\|/\left\|B^{-1}\right\|\ge\|AB\|$. **Weyls:** $\sigma_i(A+B)=\sigma_i(A)+[-\|B\|,\|B\|]$ **Rev Δ Ineq:** $\|A-B\|\ge|\|A\|-\|B\||$ **Courant Application:** $\sigma_i([A_1;A_2])\ge\max(\sigma_i(A_1),\sigma_i(A_2))$ **Schur:** Take $Av_1=\lambda_1v_1$; construct $U_1=[v_1,V_\perp]\to AU_1=U_1[e_1,X]$. Repeat. **Conditioning** $\kappa_2(A)=\sigma_1/\sigma_n=\|A\|\left\|A^{-1}\right\|$ **Similarity:** $A\to P^{-1}AP$, same λ. **Pseud-Inv:** $A^\dagger=V\Sigma^{-1}U^T=(A^TA)^{-1}A^T$. Else have $A^\dagger=A^T(AA^T)^{-1}$. $M^\dagger=M^{-1}$ if full rank.

CO: **Convex:** $f(x+\alpha(y-x))\le f(x)+\alpha(f(y)-f(x))$, or $f(y)\ge(y-x)^T\nabla f(x)+f(x)$ found via $g(t)=f(ty+(1-t)x)\le tf(y)+(1-t)f(x)$, then $g'(t)$. Other way via noting $f(y)\ge\nabla f(z)(y-z)+f(z)$, and $f(x)\ge\nabla f(z)(x-z)+f(z)$. $\times(1-t),t$ and add. **BFGS:** $\gamma_k=\Delta\nabla f_{k+1,k}=B_{k+1}(x_{k+1}-x_k)=B_{k+1}\alpha_ks_k$ **G-N:** $\vec{x}_{k+1}=\vec{x}_k-\frac{\nabla f_k}{J^TJ}$, 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:** $\|x_{k+1}-x_*\|\le(k_2(H)-1)/(k_2(H)+1)\|x_k-x_*\|$ with H hessian. Also note with EXACT linesearch for quadratic, $H(x-x_*)=-s$, and $a_k=\frac{s_k^Ts_k}{s_k^THs_k}$. **Rayleigh:** $\frac{s^THs}{\|s\|^2}\le\|H\|$ **bArm:** To show existence of α, have $\phi(\alpha)=f(x_k+\alpha_ks_k),\psi(\alpha)=\phi(\alpha)-\phi(0)-\beta\alpha\phi'(0)\le 0$, show $\psi'(0)=(1-\beta)\phi'(0)\le 0\to\psi(\alpha)\downarrow$ with α. **BFGS:** To show $H_{k+1}\ge 0$ nec. $\gamma^T\delta>0$. Suff via γ,δ LI $\to$ use $\|\cdot\|_H\to\gamma^T\delta>0$. **Quad Penalty Meth** With $y=-c/\sigma,\|\nabla_\sigma\Phi\|\le\epsilon^k,\sigma^k\to 0,x\to x_*,\nabla c(x_*)$ LI, then $y\to y_*$, $x\to KKT$, if $f,c\in C^1$. Also need J_*^T full row rank.

PROOF: If $y_* := J_*^\dagger \nabla f_* \to \|y_k - y_*\| = \left\|J_*^\dagger \nabla f_* - y_k\right\| \le \cancel{\left\|J_k^\dagger \nabla f_k - J_*^\dagger \nabla f_*\right\|} + \left\|J_k^\dagger \nabla f_k - y_k\right\|$. Next $\left\|J_k^\dagger \nabla f_k - y_k\right\| \le \left\|J_k^\dagger\right\| \|\nabla_\sigma \Phi\| \to 0$. Also, $\nabla f_* - J_*^T y_* \to 0$, and $c_{k\to *} = -\sigma^{k\to *} y_{k\to *} = 0$ so $x_* \to KKT$ **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 $\rho \ge 0.9$ then double radius, update step $x_{k+1} = x_k + s_k$. If $\rho \ge 0.1$ 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 $m_k(s_k) \le m_k(s_{kc})$, where $s_{kc} := -\alpha_{kc} \nabla f(x_k)$, and $\alpha_{kc} := \arg\min m_k(\alpha \nabla f(x_k))$ subject to $\|\alpha \nabla f\| \le \Delta$, i.e. $\alpha_{max} := \Delta / \|\nabla f\|$. **Calculation of Cauchy:** We want to prove cauchy model decrease i.e. $f(x_k) - m_k(s_k) \ge f(x_k) - m_k(s_{kc}) \ge 0.5\|\nabla f_k\| \min\left\{\Delta_k, \frac{\|\nabla f_k\|}{\|\nabla^2 f_k\|}\right\}$. 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 $H_k \le 0$ then we have $\Psi(\alpha) \le f_k - \alpha\|\nabla f_k\|^2 \to \alpha_{kc} = \alpha_{max}$. So, we have $f_k - m_{s_k} \ge f_k - m_{s_{kc}} \ge \|\nabla f_k\|\Delta_k \ge 0.5\|\nabla f_k\| \min\{\Delta_k\}$. Now B.i: If $H_k > 0 \to \alpha_{kc} = \alpha_{min}$. Here $f_k - m_{s_{kc}} = \alpha_{kc}\|\nabla f\|^2 - 0.5\alpha_{kc}^2 H_k = \frac{\|\nabla f\|^4}{2H_k} \ge \frac{\|\nabla f\|}{2} \min\left\{\frac{\|\nabla f\|}{\|\nabla^2 f\|}\right\}$ via C-S. Now B.ii: If $H_k > 0 \to \alpha_{kc} = \alpha_{max}$. Here $\Delta/\|\nabla f\| \le \|\nabla f\|^2 / H_k \to \alpha_{kc} H_k \le \|\nabla f\|^2$. So $f_k - m_{kc} = -\alpha_{kc}\|\nabla f\|^2 + \frac{\alpha_{kc}^2}{2} H_k \ge \frac{\|\nabla f\|^2}{2}\alpha_{kc} \ge 0.5\|\nabla f\| \min\{\Delta_k\}$ **TR-Global Convergence:** If $m_k(s_k) \le m_k(s_{kc})$ then either $\exists k \ge 0$ s.t. $\nabla f_k = 0$ or $\lim \|\nabla f\| \to 0$. Further, require $f \in C^2$, bounded below and also ∇f L-cont. PROOF: Using def of ρ, $f_k - f_{k+1} \ge \frac{0.1}{2}\|\nabla f_k\| \min\{\dots\}$ from above. Let $\left\|\nabla^2 f\right\| := L$, and assuming $\|\nabla f\| \ge \epsilon$ we have $f_k - f_{k+1} \ge 0.05\frac{c}{L}\epsilon^2$ assuming TR has a lower bound $c\epsilon/L$. Then sum over all successful jumps s.t. $f_0 - f_{lower} \ge \sum_{i \in \mathbb{S}} f_i - f_{i+1} \ge |\mathbb{S}| \frac{0.05 c\epsilon^2}{L}$ **Solving TR Prob:** Solve secular $\|s\|^{-1} - \Delta^{-1} = 0$. **KKT Feasibility:** Need $s^T J \ge 0$, $J_E^T s = 0$, and $s^T \nabla f < 0$. **KKT Conditions:** REMEMBER $c \ge 0$, $\lambda \ge 0$! **First Order KKT (Equality):** If we have x_* local min, then let $x = x_* + \alpha s$. Then we have $c_i(x(\alpha)) \to 0 = c_i(x_*) + \alpha s^T J \to s^T J = 0$. Further, we have $f(x) = f(x_*) + \alpha s^T \nabla f \to \alpha s^T \nabla f \ge 0$. 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 $c(x) \ge 0$, 2^{nd} order conditions are that $s^T \nabla^2 \mathcal{L} s \ge 0$ 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 $s^T J_i \ge 0 \ \forall\ i$ 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 $f \ge f(\hat{x}) + \nabla f^T(x - \hat{x})$ so $f \ge f(\hat{x}) + \hat{y}^T A(x - \hat{x}) + \sum_{i \in I} \lambda_i J_i^T (x - \hat{x})$. Choose $Ax = b$, and note that c_i concave s.t. $\lambda_i J_i^T(\hat{x})(x - \hat{x}) \ge \lambda_i(c_i(x) - c_i(\hat{x})) = \lambda_i c_i(x) \ge 0 \to f(x) \ge f(\hat{x})$. **Log-Barrier Global Convergence:** (for $f - \sum \mu \log(c_i)$) With $f \in C^1, \lambda_{ik} = \frac{\mu_k}{c_{ik}}, \|\nabla f_u(x_k)\| \le \epsilon_k, \mu_k \to 0, x_k \to x_*$. Also, $\nabla c(x_*) LI \ \forall\ i \in \mathcal{A}$ (active constraints). Then x_* KKT and $\lambda \to \lambda_*$. 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 $\lambda = \mu / c \to 0$ so $\lambda_I = 0$ as $c_I > 0$. Next $\|\nabla f_k - J_{Ak}\lambda_{Ak}\| \le \left\|\nabla f_k - J_k^T \lambda_k\right\| + \|\lambda_I\| M_1 = \|\nabla f_{\mu k}\| \to 0$. Now $\left\|J_A^\dagger \nabla f_k - \lambda_{kA}\right\| \le \left\|J_A^\dagger\right\| \left\|\nabla f_k - J_{Ak}^T \lambda_{Ak}\right\| \to 0$. So with triangle ineq $\left\|\lambda_{kA} - J_{Ak}^\dagger \nabla f_k + J_{Ak}^\dagger \nabla f_k - \lambda_{A*}\right\| \to 0$, via cont. of ∇f and $J^\dagger$. Thus $\lambda_{Ak} \to \lambda_{A*} \ge 0$. Combine s.t. $\nabla f_k - J_{Ak}^T \lambda_{AK}$ with $k \to *$ 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 $x \to x_*$ if $\sigma \to 0$ for bounded u_k or u_k to y_* for bounded σ_k. Proof via $\|c_k\| \le \sigma_k \|y_k - y_*\| + \sigma_k \|u_k - y_*\|$. If u_k bounded then $\to 0$ as $\sigma \to 0$, else trivially if $u_k \to y_*$ 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 $\lim\min\left\{\frac{|\nabla f^T s|}{\|s\|}, \left|s^T \nabla f\right|\right\}$ to 0. In non-trivial case, via bArmijo, $f_k - f_{k+1} \ge -\beta\alpha_k s_k^T \nabla f_k = \beta\alpha_k \left|s_k^T \nabla f_k\right|$. Sum s.t. $f_0 - f_{k+1} \ge \beta \sum \alpha_k \left|s_k^T \nabla f_k\right|$, so term in sum to 0. For all k successful we then have $\alpha_k \left|s_k^T \nabla f_k\right| \ge \frac{(1-\beta)\tau}{L}\left(\frac{\left|s_k^T \nabla f_k\right|}{\|s_k\|}\right)^2 \ge 0$ so squared term to 0. For unsuccesful steps $\alpha_k \ge \alpha_0$ so no norm term. **Convergence Newton LSearch:** Need $f \in C^2$ then if H_k (hessian) bdd above and below, so $\lambda_n \le \lambda(H_k) \le \lambda_1$. So $\left|s_k^T \nabla f_k\right| \ge \lambda_1^{-1}\|\nabla f_k\|^2$. Also $\|s_k\|^2 \le \lambda_n^{-2}\|\nabla f_k\|^2$. Thus $\lim\min\left\{\lambda_n \lambda_1^{-1}\|\nabla f\|, \lambda_1^{-1}\|\nabla f\|^2\right\} \to 0$ from GLM Convergence Thm. **Newton Convergence:** $f \in C^2, \nabla f_* = 0, H_*$ nonsing, H_k L-Cont, then $x_k \to x_*$ quadratic if x_k well defined for $k \ge k_0$.