Source task and conventions

Complete parts (a)–(e), including singular landmark blocks and the specified storage calculation.

Conventional solution and proof

Let $K = X^\top X \succeq 0$, write $X = [X_0 \; X_1]$, and let the compact SVD be $X_0 = U_r \Sigma_r V_r^\top$, where the diagonal entries of Σ_r are positive. Then

$$W = V_r \Sigma_r^2 V_r^\top, \quad W^\dagger = V_r \Sigma_r^{-2} V_r^\top, \quad X_0 W^\dagger X_0^\top = U_r U_r^\top =: P.$$

(a) Every principal submatrix of K is positive semidefinite: $a^\top W a = (a, 0)^\top K (a, 0) \geq 0$. Also $W W^\dagger W = W$ and $K_{21} W^\dagger W = X_1^\top X_0 V_r V_r^\top = K_{21}$. Expanding $C W^\dagger C^\top$ therefore reproduces the first row and column blocks of K, and

$$K - \widetilde{K} = \begin{pmatrix} 0 & 0 \\ 0 & K_{22} - K_{21} W^\dagger K_{21}^\top \end{pmatrix}.$$

The squared Frobenius norm is the sum of squared entries, proving the claimed norm equality.

(b) Since $C = X^\top X_0$, the SVD calculation gives $\widetilde{K} = X^\top P X$. The matrix P is the orthogonal projector onto $\operatorname{ran} X_0$: it fixes that span and annihilates its orthogonal complement. (c) For every a, $a^\top \widetilde{K} a = \|PXa\|^2 \geq 0$, so $\widetilde{K}$ is SPSD. The residual is also SPSD, since $K - \widetilde{K} = X^\top (I - P) X$. (d) If $\operatorname{rank} K = \operatorname{rank} W$, then $\dim \operatorname{ran} X = \dim \operatorname{ran} X_0$ and inclusion of the latter in the former implies equality. Consequently $PX = X$ and $\widetilde{K} = K$.

(e) With decimal $M = 10^6$, $K = 10^3$, and eight bytes per double, $m = 20{,}000{,}000$ gives $8m^2 = 3.2 \times 10^{15}$ bytes, or 3.2 PB. Storing C and W independently requires $8(ml + l^2) = 1.6008 \times 10^{12}$ bytes, or 1.6008 TB, for $l = 10{,}000$. If the duplicated W rows inside C are not separately stored, storage is $8ml = 1.6$ TB. These are raw array counts, excluding workspaces and metadata.