Appendix A-C. Vectors and Matrices

1. **Gram-Schmidt theorem:** basis $x_1, \dots, x_n$ of a vector space $\mathcal{M}$.
   (a) Can construct an orthonormal basis $\{y_1, \dots, y_n\}$.
   (b) If $\mathcal{N}$ is a subspace, an orthonormal basis $\{v_1, \dots, v_r\}$ for $\mathcal{N}$ can be extended to an orthonormal basis $\{v_1, \dots, v_r, w_1 \dots, w_{n-r}\}$ of $\mathcal{M}$ with $w_j \in \mathcal{N}^\perp$.
   (c) Any $x \in \mathcal{M}$ can be uniquely decomposed as $x = x_0 + x_1$, $x_0 \in \mathcal{N}$, $x_1 \in \mathcal{N}^\perp$.
2. **SVD of a symmetric matrix** $A = A'$: $A = PDP'$, with orthogonal $P = [v_1 \cdots v_n]$ and diagonal $D = \mathrm{diag}(\lambda_j)$. The v_j are orthonomal eignevectors of A to eigenvalues $\lambda_1, \dots, \lambda_n$
   $\Rightarrow \mathrm{tr}(A) = \sum_j \lambda_j$.
3. **SVD of rectangular** $(n \times p)$ **matrix** X: $X = UDV'$ with
   (a) $D = \mathrm{diag}(\lambda_j)$, where λ_j^2 the eigenvalues of $X'X$ pseudo $Av = \lambda v$
   (b) $(p \times p)$ $V = [v_1 \cdots v_p]$, with v_j = orthon eigenvectors of $X'X$, i.e., $X'XV = VD^2$
   (c) $(n \times p)$ $U = [u_1 \cdots u_p]$, with u_j = orthon eigenvectors of XX', with $XX'U = UD^2$
4. **Orthogonal projections:** Def: M is perpendicular projection matrix (p.p.m.) onto $C(X)$ if: $Mv = v$ for all $v \in C(X)$ and $Mw = 0$ for all $w \perp C(X)$.
   (a) M is a p.p.m iff $M = M'$ and $MM = M$.
   (b) If M is p.p.m. onto $C(X) \Rightarrow C(M) = C(X)$.
   (c) $M = OO'$, with $O = [o_1 \cdots o_p]$ where $\{o_1, \dots, o_p\}$ is an orthonormal basis of $C(X)$.
   (d) $M = X(X'X)^{-}X'$
   (e) $\mathrm{tr}(M) = r(M) = r(X)$ (eigenvalues are 1 and 0 only)
5. **Nested subspaces:** If M_0, M are p.p.m. with $C(M_0) \subset C(M)$ then $M - M_0$ is a p.p.m. with $C(M - M_0) \subset C(M)$ and $C(M - M_0) \perp C(M_0)$ (we did not see this in class, but will need it later).

§2. Estimation

6. **Estimable function:** Def: $\lambda'\beta$ is estimable if $\lambda'\beta = \rho'X\beta$. ($\Leftrightarrow \exists\, \rho$ such that $E(\rho'Y) = \lambda\beta$).
7. **LS:** $Y = X\beta + \epsilon$, and $\epsilon \sim [0, \sigma^2 I]$. Let $\mathrm{ESS}(b) = (Y - Xb)'(Y - Xb)$.
   (a) Def: $\hat\beta$ is LS estimate if $\mathrm{ESS}(\hat\beta) = \min_\beta \mathrm{ESS}(\beta)$
   (b) $\hat\beta$ is LS $\Leftrightarrow X\hat\beta = MY$, where M is p.p.m. onto $C(X)$.
8. **LS of estimable function** $\lambda'\beta$: LS estimate $\lambda'\hat\beta = \rho'X\hat\beta = \rho'MY$ (for $\lambda' = \rho'X$).
   (a) it is unbiased, i.e., $E(\lambda'\hat\beta) = E(\rho'MY) = \lambda'\beta$
   (b) if $\epsilon \sim N(0, \sigma^2 I)$ then
      i. $X\hat\beta$ is BLUE
      ii. $\hat\beta$ is m.l.e.
      iii. $\lambda'\hat\beta \sim N(\lambda'\beta, \lambda'(X'X)^{-}\lambda\sigma^2)$. Some special cases:
         - If $(X'X)^{-1}$ exists then $\hat\beta \sim N(\beta, \sigma^2(X'X)^{-1})$ and $E\{(\hat\beta - \beta)'(\hat\beta - \beta)\} = \mathrm{tr}[\sigma^2(X'X)^{-1}]$.
         - $\hat Y = X\hat\beta \sim N(X\beta, \sigma^2 M)$.
   (c) $\hat\beta$ is LS $\Leftrightarrow X'Y = X'X\hat\beta$ (*normal equations*).

9. **Unbiased estimate of** σ^2: Let $\mathrm{MSE} = Y'(I - M)Y/(n - r)$ and assume $\epsilon \sim N(0, \sigma^2 I)$
   (a) $E(\mathrm{MSE}) = E[Y'(I - M)Y/(n - r)] = \sigma^2$
   (b) quadratic form $Y'(I - M)Y/\sigma^2 \sim \chi^2(r(I - M))$
   and $Y'(I - M)Y$ is indep of MY
10. **More Sampling distributions:** Combine the last two results to find that for estimable $\lambda\beta$,

$$t \equiv \frac{\lambda'\hat\beta - \lambda'\beta}{\sqrt{\mathrm{MSE}\,\lambda'(X'X)^{-}\lambda}} \sim t(n - r),$$

   with $r = r(M)$. Some special cases and applications:
   (a) if β_j is estimable, then $t = \frac{\hat\beta_j - \beta_j}{\sqrt{\mathrm{MSE}\, H_{jj}}} \sim t(n - r)$ with $H = (X'X)^{-}$
   (b) Let $\hat Y = X\hat\beta$, and let x_i' denote the i-th row of X. Then $t = \frac{\hat Y_i - x_i'\beta}{\sqrt{\mathrm{MSE}\, x_i' H x_i}} \sim t(n - r)$
   (c) C.I. for $\lambda'\beta$ by $p(-t_{\alpha/2} \le t \le t_{\alpha/2}) = 1 - \alpha \Rightarrow p(\dots \le \lambda'\beta \le \dots) = 1 - \alpha$
   (d) Test $H_0: \lambda'\beta = \lambda'\beta_0$ with rejection region $R = \{Y : t \ge t_{\alpha/2}$ or $t \le -t_{\alpha/2}\}$ (one-sided similarly)
11. **Generalized LS:** Consider $Y = X\beta + \epsilon$ with $\epsilon \sim [0, \sigma^2 V]$, for non-singular V.
   Def: $\hat\beta$ is a *gen LS estimator* if $(Y - X\hat\beta)'V^{-1}(Y - X\hat\beta) = \min_\beta (Y - X\hat\beta)'V^{-1}(Y - X\hat\beta)$.
   Using $V = QQ'$ (from the SVD) and left multiplication with Q^{-1} we find:
   (a) $\lambda'\beta$ is estimable $\Leftrightarrow \lambda' = \rho'X$ (same as under $\mathrm{cov}(\epsilon) = \sigma^2 I$).
   (b) For estimable $\lambda'\beta$ the gen LS estimate is $\lambda'\hat\beta = \rho'AY$ with

$$A = X(X'V^{-1}X)^{-}X'V^{-1}.$$

   (c) A is a projection onto $C(X)$
   (d) $\rho'AY$ is BLUE for $\lambda'\beta$
   (e) If $\epsilon \sim N(0, \sigma^2 V)$ then
      i. $\lambda'\hat\beta$ is m.l.e.
      ii. $\lambda'\hat\beta \sim N(\lambda'\beta, \sigma^2\lambda' H\lambda)$ with $H = (X'V^{-1}X)^{-}$ [1]
      iii. Let $\mathrm{MSE} = Y'(I - A)'V^{-1}(I - A)Y/(n - r)$.
      Similar to the LS estimator we have $\mathrm{MSE}(n - r)\sigma^2 \sim \chi^2(n - r)$ and $E(\mathrm{MSE}) = \sigma^2$.
      MSE and AY are independent.
      iv. Let $H = (X'V^{-1}X)^{-}$; similar to before: $t = \frac{\lambda'\hat\beta - \lambda'\beta}{\sqrt{\mathrm{MSE}\,\lambda' H\lambda}} \sim t(n - r)$
   (f) $\hat\beta$ is gen LS $\Leftrightarrow X'V^{-1}Y = X'V^{-1}X\hat\beta$ (*normal equations*).

§15. Canonical regression & ridge regression

12. **Regression in canonical form:** Consider a regression (full rank X) model $Y = X\beta + \epsilon$, $\epsilon \sim [0, \sigma^2 I]$ and recall the SVD, $X = UDV'$.
   (a) $Y = X\beta + \epsilon \Leftrightarrow Y = UL\gamma + \epsilon$, with $\gamma = V'\beta$;
   i.e., can write the above model as a regression on the eigenvectors $u_j\lambda_j$ (canonical form regression)
   (b) LS $\hat\gamma = D^{-1}(U'Y)$ or $\hat\gamma_j = u_j'Y/\lambda_j$, $j = 1, \dots, p$.

[1] not see this in class, but it's an easy corollary

13. **Ridge regression:** Recall 8(b)iii, and use SVD for X (item 3.)

$$E[(\hat{\beta}-\beta)'(\hat{\beta}-\beta)] = \sigma^2 \mathrm{tr}[(X'X)^{-1}] = \sigma^2 \sum_j \lambda_j^{-2}.$$

This motivates

Def: $\tilde{\beta} = (X'X + kI)^{-1}X'Y$ (*ridge regression*), replacing λ_j by $\tilde{\lambda}_j^2 = \lambda_j^2 + k$.

check #2

Baysian inference for the linear model (P. Hoff, §9)

Throughout assume $p(Y \mid \beta, \sigma^2) = N(X\beta, \sigma^2 I)$ with full rank X (regression problem).

14. **Some useful results:**

    Normal/normal model: sampling model $p(Y \mid \mu) = N(\mu, V)$; prior $p(\beta) = N(m_0, \Sigma_0)$, implies posterior $p(\beta \mid Y) = N(m_1, \Sigma_1)$ with $\Sigma_1^{-1} = \Sigma_0^{-1} + V^{-1}$ and $m_1 = \Sigma_1(\Sigma_0^{-1}m_0 + V^{-1}Y)$.

    Candidate's formula: $p(\theta_2) = p(\theta_1, \theta_2)/p(\theta_1 \mid \theta_2)$ (this is useful, for example for (β, σ^2) in $p(\beta, \sigma^2 \mid Y)$).

15. **Fixed σ^2:** sampling model $p(Y \mid \beta) = N(X\beta, \sigma^2 I)$; prior $p(\beta) = N(b_0, \Sigma_0)$; then posterior $p(\beta \mid Y) = N(b_1, \Sigma_1)$ with $\Sigma_1^{-1} = \Sigma_0^{-1} + (X'X)\frac{1}{\sigma^2}$ and $b_1 = \Sigma_1(\Sigma_0^{-1}b_0 + \frac{1}{\sigma}^2 X'Y)$

16. **Random σ^2:** sampling model $p(Y \mid \beta, \sigma)$ as before, let $\gamma = 1/\sigma^2$; and prior $p(\beta, \gamma) = \mathrm{Ga}(\gamma \mid \nu_0/2, \sigma_0^2\nu_0/2) \times N(\beta \mid b_0, \Sigma_0)$; then the posterior is characterized by

    (a) $p(\beta \mid \sigma^2, Y) = N(b_1, \Sigma_1)$, as before;

    (b) $p(\gamma \mid \beta, Y) = \mathrm{Ga}\left[(\nu_0 + n)/2, \frac{1}{2}(\nu_0\sigma_0^2 + ESS(\beta))\right]$

    Use a Gibbs sampler to alternatingly generate from (a) and (b). Averaging the generated values (β, γ) across iterations approximates $E(\beta, \gamma \mid Y)$.

17. **Random σ^2 with conjugate prior:** sampling model $p(Y \mid \beta, \sigma)$ as before; and prior $p(\beta, \gamma) = \mathrm{Ga}(\gamma \mid \nu_0/2, \sigma_0^2\nu_0/2) \times N(\beta \mid b_0, \sigma^2\tilde{\Sigma}_0)$; then the posterior is characterized by

    (a) $p(\beta \mid \sigma^2, Y) = N(b_1, \Sigma_1)$, as before;

    (b) $p(\gamma \mid Y) = \mathrm{Ga}\left[(\nu_0 + n)/2, \frac{1}{2}\left\{\nu_0\sigma_0^2 + ESS(b_1) + (b_1 - b_0)'\tilde{\Sigma}^{-1}(b_1 - b_0)\right\}\right]$

    Hoff does the same with $\tilde{\Sigma} = g(X'X)^{-1}$, in which case the expression for Σ_1, above, simplifies too.