Optimization and Computational Linear Algebra – Brett Bernstein

Recitation 6

It may be helpful if after recitation you try to re-solve these problems by yourself, and use them as additional study problems for the class.

1. When solving the least squares problem
$$\min_x \|Ax - b\|^2$$
we obtained the normal equations
$$A^T Ax = A^T b.$$
   (a) Under what conditions is $A^T A$ invertible?
   (b) If $A^T A$ is not invertible, must the normal equations still have a solution?

   Solution.

   (a) $A^T A$ is invertible when $\ker(A) = 0$ (which is equivalent to saying the columns of A are linearly independent). Note, this is also exactly when A has a left inverse, and one such left inverse is given by the pseudoinverse
$$A^+ = (A^T A)^{-1} A^T.$$
   (b) Yes. Note that $A^T Ax = A^T b$ has a solution exactly when $A^T b \in \operatorname{im}(A^T A)$. But $\operatorname{im}(A^T A) = \operatorname{im}(A^T)$ (follows from the solutions to homework questions 3.3 and 3.4), so we obtain the result.
   For a quick sketch of why $\operatorname{im}(A^T A) = \operatorname{im}(A^T)$ first note that $\operatorname{im}(A^T A) \subseteq \operatorname{im}(A^T)$ is relatively easy to see. Furthermore, $\ker(A^T A) = \ker(A)$ since $Ax = 0$ if and only if $x^T A^T Ax = 0$. This implies $\operatorname{rank}(A^T A) = \operatorname{rank}(A)$ by the fundamental theorem. But we know $\operatorname{rank}(A) = \operatorname{rank}(A^T)$ so $\dim \operatorname{im}(A^T A) = \dim \operatorname{im}(A^T)$ completing the sketch.

2. Let $b \in \mathbb{R}^n$, let $A \in \mathbb{R}^{n \times 2}$ have the form
$$A = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_n \end{bmatrix},$$
and let $x^* \in \mathbb{R}^2$ be a minimizer for the least squares problem
$$\min_x \|Ax - b\|^2.$$
Prove that $\sum_{i=1}^n (Ax^* - b)_i = 0$. [Note: In linear regression terminology, this is the statement that the residuals have zero mean.]

Solution. Since $Ax^* = P_{\text{im}(A)}(b)$ we have $Ax - b \perp \text{im}(A)$. But the all 1's vector $\mathbf{1}$ is in $\text{im}(A)$, so

$$0 = \mathbf{1}^T(Ax^* - b) = \sum_{i=1}^{n}(Ax^* - b)_i.$$

3. For any $\alpha \in \mathbb{R}^3$ define the function $f_\alpha : \mathbb{R} \to \mathbb{R}$ by $f_\alpha(x) = \alpha_1 + \alpha_2 x + \alpha_3 x^2$. Given a dataset $(t_1, y_1), \ldots, (t_n, y_n) \in \mathbb{R}^2$ show how to solve the least squares problem

$$\min_{\alpha} \sum_{i=1}^{n}(f_\alpha(t_i) - y_i)^2.$$

Solution. Define a matrix $A \in \mathbb{R}^{n\times 3}$ by

$$A = \begin{bmatrix} 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \\ \vdots & \vdots & \vdots \\ 1 & t_n & t_n^2 \end{bmatrix}.$$

Then we have

$$\sum_{i=1}^{n}(f_\alpha(t_i) - y_i)^2 = \|A\alpha - y\|^2,$$

where $y \in \mathbb{R}^n$ has y_i as its ith component. Thus we can obtain the solution by solving the normal equations

$$A^T A\alpha = A^T y$$

for α.

4. Let $A \in \mathbb{R}^{m\times n}$ have linearly independent columns.

   (a) Give an expression for $P_{\text{im}(A)}$ in terms of the QR-factorization of A.

   (b) Give an expression for $P_{\text{im}(A)}$ in terms of A.

   Solution.

   (a) The QR factorization lets us write $A = QR$ where Q has orthonormal columns and R is upper triangular. Here the columns of Q form an orthonormal basis for $\text{im}(A)$ so we have $P_{\text{im}(A)} = QQ^T$ by the previous lab.

   (b) If x^* is a minimizer for $\|Ax - b\|^2$ then $Ax^* = P_{\text{im}(A)}(b)$. Thus we have

$$P_{\text{im}(A)} = A(A^T A)^{-1}A^T = AA^+.$$

5. If Q is orthogonal prove that $\langle x, y\rangle = \langle Qx, Qy\rangle$ for all $x, y \in \mathbb{R}^n$.

Solution. Note that

$$\langle Qx, Qy\rangle = (Qx)^T(Qy) = x^T Q^T Q y = x^T y = \langle x, y\rangle.$$

6. Let U be a subspace of $\mathbb{R}^n$. Show that $\dim(U) + \dim(U^\perp) = n$.

   Solution. Let $v_1, \ldots, v_k$ be a basis for U and let $w_1, \ldots, w_p$ be a basis for $U^\perp$. We claim that $v_1, \ldots, v_k, w_1, \ldots, w_p$ is a basis for $\mathbb{R}^n$ (so that $k + p = n$). First note that the list is linearly independent by a result from lab 2 since $U \cap U^\perp = \{0\}$: if $v \in U \cap U^\perp$ then $\langle v, v\rangle = 0$ so $v = 0$. We know that our list spans $\mathbb{R}^n$ since homework 5 shows that any vector $u \in \mathbb{R}^n$ can be written as $u = x + y$ where $x \in U$ and $y \in U^\perp$.

7. Consider the block matrices

   $$J = \begin{bmatrix} A & B \end{bmatrix} \in \mathbb{R}^{m\times(n+k)} \quad \text{and} \quad Q = \begin{bmatrix} C \\ D \end{bmatrix} \in \mathbb{R}^{(n+k)\times p}$$

   where $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{m\times k}$, $C \in \mathbb{R}^{n\times p}$ and $D \in \mathbb{R}^{k\times p}$. Compute JQ and Qx where $x \in \mathbb{R}^p$.

   Solution. $JQ = AC + BD$ and $Qx = \begin{bmatrix} Cx \\ Dx \end{bmatrix}$.

8. Let $f, g : [0, 1] \to \mathbb{R}$ be defined by $f(x) = 1$ and $g(x) = x$. Recall that the continuous functions from $[0, 1]$ to $\mathbb{R}$ form a vector space with inner product

   $$\langle p, q\rangle = \int_0^1 p(t)q(t)\,dt.$$

   (a) Show that f, g are linearly independent.

   (b) We just showed that f, g is a basis for $\mathrm{Span}(f, g)$. Call this basis B. Define the *coordinatization* function $\Phi_B : \mathrm{Span}(f, g) \to \mathbb{R}^3$ by

   $$\Phi_B(\alpha_1 f + \alpha_2 g) = \alpha.$$

   That is, it maps a vector in $\mathrm{Span}(f, g)$ to its vector of coefficients. Compute $\Phi_B(j)$ where $j(x) = 4x - 3$.

   (c) Compute an orthonormal basis C for $\mathrm{Span}(f, g)$.

   (d) Compute $\Phi_C(j)$ where $j(x) = 4x - 3$.

   (e) Let $k(x) = x^2$. Compute $P_{\mathrm{Span}(f,g)}(k)$.

   (f) In what sense is $P_{\mathrm{Span}(f,g)}(k)$ the best approximation of k.

   Solution.

   (a) Note that g is not constant, so it is not a scalar multiple of f.

(b) $\Phi_B(j) = (-3, 4)$.

(c) We apply Gram-Schmidt. Let $u_1 = f$ since $\|f\| = 1$. Define $w_2 : [0,1] \to \mathbb{R}$ by

$$w_2 = g - \langle g, u_1\rangle u_1 = x - \int_0^1 t\,dt = x - 1/2.$$

Then we set $u_2 = w_2/\|w_2\|$ where

$$\|w_2\|^2 = \int_0^1 (x-1/2)^2\,dx = \int_{-1/2}^{1/2} u^2\,du = \frac{1}{12}$$

so $u_2 = \sqrt{12}(x - 1/2)$.

(d) $\Phi_C(j) = (-1, 4/\sqrt{12})$

(e) Recall that

$$P_{\mathrm{Span}(f,g)}(k) = \langle k, u_1\rangle u_1 + \langle k, u_2\rangle u_2.$$

Computing we have

$$\begin{array}{rcl} \langle k, u_1\rangle & = & \int_0^1 x^2\,dx = 1/3 \\ \langle k, u_2\rangle & = & \sqrt{12}\int_0^1 x^3 - x^2/2\,dx = \sqrt{12}(1/4 - 1/6) = \frac{1}{\sqrt{12}}. \end{array}$$

Thus

$$P_{\mathrm{Span}(f,g)}(k) = 1/3 + (x - 1/2) = x - 1/6.$$

(f) By definition, $P_{\mathrm{Span}(f,g)}(k)$ is the minimizer of

$$\min_{h\in\mathrm{Span}(f,g)} \|h - k\|^2$$

where

$$\|h-k\|^2 = \int_0^1 (h(x) - k(x))^2\,dx.$$

9. We can also use coordinatization for $\mathbb{R}^n$. If we have a basis $B = v_1, \ldots, v_n$ for $\mathbb{R}^n$ then we can define the coordinatization (or change-of-basis) map $\Phi_B : \mathbb{R}^n \to \mathbb{R}^n$ by

$$\Phi_B(\alpha_1 v_1 + \cdots + \alpha_n v_n) = \alpha.$$

(a) Let B denote the basis $(1,0), (-1,1)$ for $\mathbb{R}^2$. Compute

$$\Phi_B\left(\begin{bmatrix}1\\0\end{bmatrix}\right), \quad \Phi_B\left(\begin{bmatrix}-1\\1\end{bmatrix}\right), \quad \text{and} \quad \Phi_B\left(\begin{bmatrix}0\\1\end{bmatrix}\right).$$

(b) Suppose $B = v_1, \ldots, v_n$ is a basis for $\mathbb{R}^n$. Give the matrices corresponding to Φ_B and Φ_B^{-1} (possible since $\Phi_B : \mathbb{R}^n \to \mathbb{R}^n$ is linear and invertible).

(c) For which bases B of $\mathbb{R}^n$ does Φ_B preserve inner products? That is, for which bases B does

$$\langle \Phi_B(x), \Phi_B(y) \rangle = \langle x, y \rangle$$

for all $x, y \in \mathbb{R}^n$?

Solution.

(a)

$$\begin{aligned} \Phi_B\left(\begin{bmatrix}1\\0\end{bmatrix}\right) &= \begin{bmatrix}1\\0\end{bmatrix}, \\ \Phi_B\left(\begin{bmatrix}-1\\1\end{bmatrix}\right) &= \begin{bmatrix}0\\1\end{bmatrix}, \\ \Phi_B\left(\begin{bmatrix}0\\1\end{bmatrix}\right) &= \begin{bmatrix}1\\1\end{bmatrix}. \end{aligned}$$

(b) Let $A \in \mathbb{R}^{n\times n}$ denote the matrix with v_i as its ith column. Then $\Phi_B = A^{-1}$ and $\Phi_B^{-1} = A$.

(c) Orthonormal bases (see question 5 above).