Optimization and Computational Linear Algebra – Brett Bernstein

Recitation 5

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. Compute $\|ax\|$ for $a \in \mathbb{R}$ and $x \in \mathbb{R}^n$.

   Solution. Below we show that $\|ax\| = |a|\|x\|$.

   $$\|ax\| = \sqrt{\langle ax, ax\rangle} = \sqrt{a^2\langle x, x\rangle} = |a|\|x\|.$$

2. When does $\|x + y\| = \|x\| + \|y\|$ for $x, y \in \mathbb{R}^n$?

   Solution. This requires $(\|x\| + \|y\|)^2 = \|x + y\|^2$ which occurs exactly when $\|x\|\|y\| = x^T y$. This happens when when $y = 0$ or when $x = ay$ for some $a \geq 0$. To see this, either check when equality occurs in Cauchy-Schwarz, or note that we need $\theta = 0$. For the Cauchy-Schwarz method, we showed that for $x \neq 0$ and $y \neq 0$ we have

   $$0 \leq \left\| \frac{x}{\|x\|} - \frac{y}{\|y\|} \right\|^2 = 2 - 2\frac{\langle x, y\rangle}{\|x\|\|y\|} \iff \langle x, y\rangle \leq \|x\|\|y\|.$$

   But the first inequality is an equality whenever x is a positive multiple of y.

3. Prove the parallelogram identity holds for any $x, y \in \mathbb{R}^n$:

   $$2\|x\|^2 + 2\|y\|^2 = \|x + y\|^2 + \|x - y\|^2.$$

   Solution. Note that

   $$\|x+y\|^2+\|x-y\|^2 = (\|x\|^2+\|y\|^2+2\langle x, y\rangle)+(\|x\|^2+\|y\|^2-2\langle x, y\rangle) = 2\|x\|^2+2\|y\|^2.$$

4. ($\star$) For any $A \in \mathbb{R}^{m\times n}$ and $x \in \mathbb{R}^n$ show that

   $$\|Ax\| \leq \|x\|\sqrt{\sum_{i=1}^m \sum_{j=1}^n A_{ij}^2}.$$

   Solution. Letting a_i^T denote the ith row of A we have

   $$\begin{array}{rcll} \|Ax\|^2 & = & (a_1^T x)^2 + \cdots + (a_m^T x)^2 & (a_i^T x = (Ax)_i) \\ & \leq & \|a_1\|^2\|x\|^2 + \cdots + \|a_m\|^2\|x\|^2 & \text{(Cauchy-Schwarz)} \\ & = & \|x\|^2(\|a_1\|^2 + \cdots + \|a_m\|^2) & \\ & = & \|x\|^2 \sum_{i=1}^m \sum_{j=1}^n A_{ij}^2. & \end{array}$$

5. Let $v_1, \ldots, v_m \in \mathbb{R}^n$ be linearly independent. Show there is an orthonormal basis for $\mathrm{Span}(v_1, \ldots, v_m)$.

   Solution. We will outline an algorithm known as Gram-Schmidt.

   (a) Set $w_1 = v_1$ and $u_1 = w_1/\|w_1\|$.

   (b) For $i = 2, \ldots, m$:

      i. Define w_i by
      $$w_i = v_i - \langle v_i, u_1\rangle u_1 - \cdots - \langle v_i, u_{i-1}\rangle u_{i-1}$$

      ii. Let $u_i = w_i/\|w_i\|$.

   We claim that $u_1, \ldots, u_m$ are orthonormal and that $u_i \in \mathrm{Span}(v_1, \ldots, v_i)$ for all i. The claim implies $\mathrm{Span}(u_1, \ldots, u_i) \subseteq \mathrm{Span}(v_1, \ldots, v_i)$ with both spans having dimension i for all $i = 1, \ldots, m$. This shows the spans are equal and completes the proof.

   Proof of claim. Proof by induction. More precisely, we show that for all $i \geq 1$ we have $\langle u_i, u_j\rangle = 0$ for any $j < i$, $\langle u_i, u_i\rangle = 1$, and $u_i \in \mathrm{Span}(v_1, \ldots, v_i)$. For the base case $i = 1$ we only need that $v_1 \neq 0$ (so that u_1 is well-defined), but this is immediate from linear independence. For the induction case, assume the statement holds up to $i \geq 1$. By the definition of w_{i+1} and the induction hypothesis we have
   $$w_{i+1} \in \mathrm{Span}(v_{i+1}, u_1, \ldots, u_i) \subseteq \mathrm{Span}(v_{i+1}, v_1, \ldots, v_i).$$
   If $w_{i+1} = 0$ then $v_{i+1} \in \mathrm{Span}(v_1, \ldots, v_i)$ contradicting linear independence. Thus $w_{i+1} \neq 0$, u_{i+1} is well-defined, and $\|u_{i+1}\| = 1$. Since $u_{i+1} = w_{i+1}/\|w_{i+1}\|$ we also have
   $$u_{i+1} \in \mathrm{Span}(v_{i+1}, v_1, \ldots, v_i).$$
   Furthermore, for any $j < i + 1$ we have
   $$\begin{aligned} \|w_{i+1}\|\langle u_{i+1}, u_j\rangle &= \langle w_{i+1}, u_j\rangle \\ &= \langle v_{i+1} - \textstyle\sum_{k=1}^{i}\langle v_{i+1}, u_k\rangle u_k, u_j\rangle \\ &= \langle v_{i+1}, u_j\rangle - \textstyle\sum_{k=1}^{i}\langle v_{i+1}, u_k\rangle\langle u_k, u_j\rangle \\ &= \langle v_{i+1}, u_j\rangle - \langle v_{i+1}, u_j\rangle \qquad \text{(Induction Hypothesis)} \\ &= 0. \end{aligned}$$
   □

6. What is the output of Gram-Schmidt if the input vectors $v_1, \ldots, v_m$ are already orthonormal?

   Solution. It simply sets $u_i = v_i$.

7. Let $v_1, \ldots, v_m \in \mathbb{R}^n$ be orthonormal. Prove that $v_1, \ldots, v_m$ can be extended (if necessary) to form an orthonormal basis for $\mathbb{R}^n$.

Solution. Since $v_1, \dots, v_m$ are linearly independent, we can extend them to form a basis for $\mathbb{R}^n$. Running Gram-Schmidt on the new basis leaves $v_1, \dots, v_m$ unchanged, and modifies the remaining vectors to form an orthonormal basis.

8. Let $A \in \mathbb{R}^{m\times n}$ have linearly independent columns. Show there is a matrix $Q \in \mathbb{R}^{m\times n}$ and $R \in \mathbb{R}^{n\times n}$ such that $A = QR$, Q has orthonormal columns, and R is upper triangular.

   Solution. Run Gram-Schmidt on the columns of A to obtain orthonormal vectors $u_1, \dots, u_n \in \mathbb{R}^m$. Note that the Gram-Schmidt algorithm ensures that $v_k \in \mathrm{Span}(u_1, \dots, u_k)$, where in this case v_i denotes the ith column of A (to see this, note that above we proved that $\mathrm{Span}(u_1, \dots, u_i) = \mathrm{Span}(v_1, \dots, v_i)$ for $i = 1, \dots, m$). But saying $v_k \in \mathrm{Span}(u_1, \dots, u_k)$ is exactly the statement that $A = QR$ for some upper triangular R (think about column method of multiplication).

9. Let U be a subspace of $\mathbb{R}^n$ with orthonormal basis $u_1, \dots, u_k$ which we extend to an orthonormal basis $u_1, \dots, u_k, u_{k+1}, \dots, u_n$. Show how to compute $P_U(v)$ for any $v \in \mathbb{R}^n$ using this basis.

   Solution. Recall that $P_U(v) = \arg\min_{x\in U} \|v - x\|$. Writing v and x in terms of the orthonormal basis we have

$$v = \alpha_1 u_1 + \cdots + \alpha_k u_k + \alpha_{k+1} u_{k+1} + \cdots + \alpha_n u_n$$

   and

$$x = \beta_1 u_1 + \cdots + \beta_k u_k.$$

   Thus, by the homework 5 question 2, $\|v - x\|$ is minimized when $\beta_i = \alpha_i$ for $i = 1, \dots, k$. Thus we have

$$P_U(v) = \alpha_1 u_1 + \cdots + \alpha_k u_k.$$

   Stated simply, we only keep the terms from the subspace basis and omit the remaining ones.

10. Let U be a subspace of $\mathbb{R}^n$ and let $v \in \mathbb{R}^n$. Prove that $\langle v - P_U(v), x\rangle = 0$ for any $x \in U$.

    Solution. Choose an orthonormal basis $u_1, \dots, u_k$ for U as in the previous problem, and extend it to form an orthonormal basis $u_1, \dots, u_k, u_{k+1}, \dots, u_n$ for $\mathbb{R}^n$. Then if v is given by

$$v = \alpha_1 u_1 + \cdots + \alpha_n u_n$$

    then

$$v - P_U(v) = \alpha_{k+1} u_{k+1} + \cdots + \alpha_n u_n.$$

    Writing $x \in U$ as

$$x = \beta_1 u_1 + \cdots + \beta_k u_k$$

we have

$$\langle v - P_U(v), x\rangle = \langle \alpha_{k+1}u_{k+1} + \cdots + \alpha_n u_n, \beta_1 u_1 + \cdots + \beta_k u_k\rangle = 0.$$

11. Let U be a subspace of $\mathbb{R}^n$ with orthonormal basis $u_1, \ldots, u_k$. Give the matrix corresponding to the projection P_U in terms of $u_1, \ldots, u_k$.

Solution. Let Q denote the matrix with $u_1, \ldots, u_k$ as columns. Extend the basis $u_1, \ldots, u_k$ to an orthonormal basis $u_1, \ldots, u_n$ for $\mathbb{R}^n$. Then QQ^T satisfies $QQ^T u_i = u_i$ for $i = 1, \ldots, k$ and $QQ^T u_j = 0$ for $j = k+1, \ldots, n$. To see this note that $Q^T u_i = e_i$ for $i = 1, \ldots, k$ and $Q^T u_j = 0$ for $j > k$. Thus if $v \in \mathbb{R}^n$ is given by

$$v = \alpha_1 u_1 + \cdots + \alpha_n u_n$$

then

$$QQ^T v = \alpha_1 u_1 + \cdots + \alpha_k u_k = P_U(v)$$

as required.