HOMEWORK FOR 18.101, FALL 2007
ASSIGNMENT 1 SOLUTIONS

(1) Given a linear map

$$T : \mathbb{R}^m \longrightarrow \mathbb{R}^n$$

Define the operator norm of T as follows:

$$\|T\| := \sup_{x \neq 0} \frac{\|T(x)\|}{\|x\|}$$

Similarly, if A is a matrix, define the operator norm of A by

$$\|A\| := \sup_{x \neq 0} \frac{\|Ax\|}{\|x\|}$$

(a) Show that $\|T\|$ is finite.
(b) Show that $\|\cdot\|$ is a norm.
(c) Show that if AB is defined,

$$\|A\| \, \|B\| \geq \|AB\|$$

Solution:

(a) By rescaling x to $\frac{x}{\|x\|}$ (which has norm equal to 1), and by using the fact that T is linear, we get an alternative characterization:

$$\|T\| = \sup_{\|x\|=1} \|T(x)\|$$

We will now show that this quantity is finite. Let us think of T as an $m \times n$ matrix in the corresponding standard bases of $\mathbb{R}^n$ and $\mathbb{R}^m$. We find a constant $M > 0$ such that all the entries of T are bounded in absolute value by M. Let $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ be such that $\|x\| = 1$. We know that then $|x_j| \leq 1 \forall j = 1, \ldots, n$. Using the triangle inequality and the bounds on the entries of T, it follows that every entry of $T(x)$ is bounded in absolute value by nM. Hence $\|T(x)\| \leq nM\sqrt{n}$. We can take supremums over all $x \in \mathbb{R}^n$ such that $\|x\| = 1$ to obtain that: $\|T\| \leq nM\sqrt{n} < +\infty$ Hence, $\|T\|$ is indeed finite.

(b) In order to show that $\|\cdot\|$ indeed gives us a norm on the space $L(\mathbb{R}^n, \mathbb{R}^m)$ of linear operators from $\mathbb{R}^n$ to $\mathbb{R}^m$, we need to check the following properties:

(i) **Positive Definiteness:** For all $T \in L(\mathbb{R}^n, \mathbb{R}^m)$, $\|T\| \geq 0$ and $\|T\| = 0$ if and only if $T = 0$.
The fact that $\|T\| \geq 0, \ \forall T \in L(\mathbb{R}^n, \mathbb{R}^m)$ follows from definition of $\|\cdot\|$. It is also an immediate consequence of the definition that $\|0\| = 0$. Suppose $T \in L(\mathbb{R}^n, \mathbb{R}^m)$ is such that $\|T\| = 0$. Then $T(x) = 0, \ \forall x \in \mathbb{R}^n$ such that $\|x\| = 1$. Since every element of $\mathbb{R}^n$ is a scalar multiple of an element of unit norm, it follows by linearity of T that $T = 0$ on $\mathbb{R}^n$.

(ii) **Homogeneity:** Let $T \in L(\mathbb{R}^n, \mathbb{R}^m)$ and $\lambda \in \mathbb{R}$ be given. We must show that:$\|\lambda T\| = |\lambda| \|T\|$.
To see this, suppose that $x \in \mathbb{R}^n$ and $\|x\| = 1$. Then $\|(\lambda T)(x)\| = |\lambda| \|T(x)\|$. Taking supremums over $x \in \mathbb{R}^n$ with $\|x\| = 1$, it follows that $\|\lambda T\| = |\lambda| \|T\|$.

(iii) **Triangle Inequality:** We must show that $\forall A, B \in L(\mathbb{R}^n, \mathbb{R}^m)$ $\|A + B\| \leq \|A\| + \|B\|$. To see this, let A, B be as above. Let $x \in \mathbb{R}^n$ be such that $\|x\| = 1$. Then, by the triangle inequality in $\mathbb{R}^n$, we have that:

$$\|(A + B)(x)\| \leq \|A(x)\| + \|B(x)\|$$

Taking supremums over $\|x\| = 1$, we indeed get:

$$\|A + B\| \leq \|A\| + \|B\|$$

(c) We observe that $\forall T \in L(\mathbb{R}^n, \mathbb{R}^m)$, $\forall x \in \mathbb{R}^n$, we have: $\|T(x)\| \leq \|T\| \|x\| (*)$. Namely, this fact evidently holds in the case that $x = 0$ and the case $x \neq 0$ follows from the original definition of $\|T\|$. Suppose now that A and B are such that AB is well defined. Let v be an element of the domain of B and suppose that $\|v\| = 1$. We obtain that: $\|(AB)(v)\| = \|A(B(v))\| \leq \{\text{by using } (*)\} \leq \|A\| \|B(v)\| \leq \{\text{by using } (*) \text{ again }\} \leq \|A\| \|B\| \|v\|$. Taking supremums over $\|v\| = 1$, it follows that indeed $\|A\| \|B\| \geq \|AB\|$, as was claimed.

(2) Prove that the two norms $|\cdot|_s$ and $\|\cdot\|$ on $\mathbb{R}^n$ give the same topology in the sense that if U is an open set using the metric from one norm, it is open using the metric from the other norm.
(Recall that we defined $|x|_s := \max |x_i|$, and $\|x\| := \sqrt{\sum x_i^2}$.)
Solution:
We observe that $\forall x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ we have: $|x|_s \leq \|x\|$. This is because for $j = 1, \ldots, n$ we have: $|x_j| \leq \sqrt{\sum x_i^2}$.
On the other hand, $|x_j| \leq |x|_s \, \forall j = 1, \ldots, n$. We thus get that: $\|x\| = \sqrt{\sum x_i^2} \leq \sqrt{n} \, |x|_s$.
From the preceding discussion, we deduce that:

$$\forall x \in \mathbb{R}^n, \; |x|_s \leq \|x\| \leq \sqrt{n} \, |x|_s \qquad (*)$$

Using $(*)$, we now prove the claim.
Let us now define the notation that we will use for the rest of the problem. Let $p \in \mathbb{R}^n$, and $r > 0$ be given. We denote by

$$B_r(p) := \{x \in \mathbb{R}^n, \; \|x - p\| < r\}$$

$$B_{r,s}(p) := \{x \in \mathbb{R}^n, \; |x - p|_s < r\}$$

- Suppose that $U \subseteq \mathbb{R}^n$ is open with respect to $\|\cdot\|$. If U is empty, then U is automatically open with respect to $|\cdot|_s$. Hence, it suffices to consider the case when U is nonempty. Let us choose $p \in U$. Then, since U is open with respect to $\|\cdot\|$, we can find $r > 0$ such that $B_r(p) \subseteq U$. From (*), we have that $B_{\frac{r}{\sqrt{n}},s}(p) \subseteq B_r(p)$. Hence $B_{\frac{r}{\sqrt{n}},s}(p) \subseteq U$, so U is open with respect to $|\cdot|_s$.

- Suppose that $U \subseteq \mathbb{R}^n$ is open with respect to $|\cdot|_s$. As before, we consider the case of nonempty U and let $p \in U$ be given. We can find $r > 0$ such that $B_{r,s}(p) \subseteq U$. By using (*), we know that $B_r(p) \subseteq B_{r,s}(p)$.Hence $B_r(p) \subseteq U$,so U is open with respect to $\|\cdot\|$.

  Conclusion: $\|\cdot\|$ and $|\cdot|_s$ define the same topology.

(3) (a) Show that given any $m \times n$ matrix A, the transpose of the matrix A^T is the unique $n \times m$ matrix with the property that

$$(Ax) \cdot y = x \cdot (A^T)y \quad \forall x \in \mathbb{R}^n, y \in \mathbb{R}^m$$

(b) Show that

$$\|A\| = \left\|A^T\right\|$$

Solution:

(a) Suppose that A is an $m \times n$ matrix. Let $\{e_j \ , j = 1, \dots, n\}, \{f_i \ , i = 1, \dots, m\}$ denote the standard bases for $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively. We know from bilinearity that an $n \times m$ matrix B satisfies $(Ax) \cdot y = x \cdot (By)$ $\forall x \in \mathbb{R}^n, y \in \mathbb{R}^m$ if and only if we have that $\forall i = 1, \dots, m, j = 1, \dots, n$ $(Ae_j) \cdot f_i = e_j \cdot (Bf_i)$ Now, $\forall i = 1, \dots, m, j = 1, \dots, n$, we have that: $(Ae_j) \cdot f_i = i^{th}$ entry of the j^{th} column of A, which we denote by $A_{i\,j}$. Also, $e_j \cdot (Bf_i) = j^{th}$ entry of the i^{th} column of B, which we denote by $B_{j\,i}$. Hence, from the above we may conclude that, given an $n \times m$ matrix B we have that:

$$(Ax) \cdot y = x \cdot (By) \ \forall x \in \mathbb{R}^n, y \in \mathbb{R}^m$$

if and only if

$$A_{i\,j} = B_{j\,i} \, \forall i = 1, \dots, m, j = 1, \dots, n.$$

The latter set of equalities is equivalent to the fact that $B = A^T$, i.e. that B is the *transpose* of A.

(b) Let us show that $\left\|A^T\right\| \leq \|A\|$. Since $(A^T)^T = A$, i.e since $(\cdot)^T$ is an *involution*, the claim will then follow. We know from Problem 1 that, given any linear map $B : \mathbb{R}^p \longrightarrow, \mathbb{R}^q$ $\|B\| = \sup_{\|v\|=1} \|Bv\|$.

Suppose now that $v \in \mathbb{R}^n$ with $\|v\| = 1$ and $w \in \mathbb{R}^m$ with $\|w\| = 1$ are given.

Then we know that $(A^T)w \cdot v = w \cdot Av$. Combining the Schwarz-Cauchy Inequality and the fact that $\|Av\| \leq \|A\| \|v\|$, it follows that the right hand side of the above inequality has absolute value less than or equal to: $\|w\| \|A\| \|v\| = \|A\|$.

In particular, given $w \in \mathbb{R}^m$ with $\|w\| = 1$, we can find $v \in \mathbb{R}^n$ with $\|v\| = 1$ such that $(A^T)w \cdot v = \left\|(A^T)w\right\|$ Namely, we let $v := \frac{(A^T)w}{\|(A^T)w\|}$ if $(A^T)w \neq 0$ and we let v be an arbitrary element of unit norm in $\mathbb{R}^m$ if $(A^T)w = 0$.

It follows from the previous discussion that, $\forall w \in \mathbb{R}^m$ with $\|w\| = 1$ we have $\left\|(A^T)w\right\| \leq \|A\|$ Taking supremums over such w, it follows that:

$$\left\|A^T\right\| \leq \|A\|$$

As we noted earlier, from this inequality we may deduce that:

$$\left\|A^T\right\| = \|A\|$$

(4) (a) Suppose that $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is differentiable, and suppose that for all x, $\|Df(x)\| \leq 1$. Prove that

$$\|f(x) - f(y)\| \leq \|x - y\|$$

(Hint: Try taking the dot product with $f(x) - f(y)$, and use the chain rule to convert this into a single variable problem. Then you can use the mean value theorem.)

(b) Find a counterexample to the following naive generalization of the mean value theorem: Given $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ differentiable and points $x, y \in \mathbb{R}^n$, there exists some point c on the line segment between x and y so that

$$Df(c)(x - y) = f(x) - f(y)$$

Solution:

(a) Let $x, y \in \mathbb{R}^n$ be given. Let us define:
$g : \mathbb{R} \longrightarrow \mathbb{R}$ by
$g(t) := (f(y+t(x-y))-f(y))\cdot(f(x)-f(y))$. We observe that then g is a differentiable function of t. Also $g(0) = 0$ and $g(1) = (\|f(x) - f(y)\|)^2$
By applying the Chain Rule, we know that $\forall t \in \mathbb{R}$

$$g'(t) = (Df(y + t(x - y))(x - y)) \cdot (f(x) - f(y))$$

We can now use the Mean Value Theorem to get that there exists some $t_0 \in (0, 1)$ so that

$$g(1) - g(0) = g'(t_0)$$

Therefore,

$$\|f(x) - f(y)\|^2 = (Df(y + t_0(x - y))(x - y)) \cdot (f(x) - f(y))$$

Now we can use the Cauchy-Schwartz inequality

$$\|f(x) - f(y)\|^2 \leq \|Df(y + t_0(x - y))(x - y)\| \, \|f(x) - f(y)\|$$

Therefore, as all terms are non negative,

$$\|f(x) - f(y)\| \leq \|Df(y + t_0(x - y))(x - y)\| \leq \|Df(y + t_0(x - y))\| \, \|x - y\| \leq \|x - y\|$$

(b) We consider the function $f : \mathbb{R} \longrightarrow \mathbb{R}^2$ be such that $f(t)$ is:

$$\begin{bmatrix} \cos t \\ \sin t \end{bmatrix}$$

Then f is differentiable and $Df(t)$ is given by

$$\begin{bmatrix} -\sin t \\ \cos t \end{bmatrix}$$

Since $(\sin t)^2 + (\cos t)^2 = 1$, it follows that $Df(t)$ has rank $1 \forall t \in \mathbb{R}$, so $Df(t)$ is injective $\forall t \in \mathbb{R}$. Observe that then $f(0) = f(2\pi)$. However, if there existed a point c on the line segment joining 0 and 2π such that $Df(c)(2\pi - 0) = f(2\pi) - f(0)$, then it would follow that $Df(c)(2\pi) = 0$. This is impossible since $2\pi \neq 0$ and since $Df(c)$ is injective. Thus, the 'naive' generalization of the Mean Value Theorem doesn't hold.

(5) (a) Let $f : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ be of class C^1. Prove that the set $S \subset \mathbb{R}^n$ consisting of points $x \in \mathbb{R}^n$ where $Df(x)$ has rank n is open.
(Hint: The determinant has a formula which is a polynomial in the coefficients of the matrix. This tells you that the determinant in a continuous function of the coefficients of a matrix. Use this.)
(b) Use the inverse function theorem to prove that $f(S) \subset \mathbb{R}^n$ is also open.

Solution:

(a) We observe that:

$$S = \{x \in \mathbb{R}^n,\ \ Df(x) \text{ has rank n}\} = \{x \in \mathbb{R}^n,\ \ \det(Df(x)) \neq 0\}$$

We also know that the determinant of a matrix is a polynomial in its entries. This follows from the definition of a determinant. Now, for f as above, the quantity $\det(Df(x))$ is a polynomial in the n^2 quantities $\frac{\partial f_j}{\partial x_i}(x)$, where $i, j = 1, \ldots, n$. Since f is a class C^1 function, it follows that $\frac{\partial f_j}{\partial x_i}(x)$ is a continuous function in x for $i, j = 1, \ldots, n$. From here we may deduce that $\det(Df(x))$ is a continuous function in $x \in \mathbb{R}^n$. Hence, by continuity, the set $S = \{x \in \mathbb{R}^n,\ \ \det(Df(x)) \neq 0\}$ is open as it is the inverse image of an open subset of $\mathbb{R}$.

(b) In order to prove the claim, we want to use the Inverse Function Theorem.
Suppose that $y \in f(S)$ is given. By definition, we can find $x \in S$ such that $f(x) = y$. Then, since $x \in S$, it follows that $\det(Df(x)) \neq 0$ so by the Inverse Function Theorem, we can find a neighborhood U of x in $\mathbb{R}^n$ and a neighborhood V of y in $\mathbb{R}^n$ such that f is a bijection from U to V which has an inverse that is of class C^1. Call this inverse g. In particular, we know by the Chain Rule that:
$(Df(w))\circ(Dg(f(w))) = I_n\ \forall w \in U$. Hence $Df(w)$ is invertible $\forall w \in U$. Thus, it follows that $U \subseteq S$. Hence, $V = f(U) \subseteq f(S)$. V is also a neighborhood of y in $\mathbb{R}^n$ by construction. Since such a neighborhood V can be found for all $y \in f(S)$, it follows that $f(S)$ is open, as was claimed.