1 Sequences, Limits, Functions

1. Let $s = \{x_i\}$ be a non-decreasing sequence, bounded from above. By the least upper-bound property, s has a least upper-bound. Call it u. Pick $\epsilon > 0$. Then $\exists k$ such that $x_i > u - \epsilon$ for some $i > k$. If not, then $u - \epsilon$ would be an upper-bound of s, contradicting our choice of u as the least upper-bound. But if $x_i > u - \epsilon$, then so is x_j, for all $j > i$, because s is non-decreasing. It follows that s converges to u.

2. For some $k > 0$, it must be the case that $u_i \geq v_i$, $\forall i > k$. If not, then $\exists i$ such that $u_i < v_i$. But then $u_j < v_j$, and $v_{j+1} - u_{j+1} \geq v_j - u_j$, $\forall j > i$, contradicting the assumption that $\lim u_n - v_n = 0$. So $\{v_n\}$ is bounded above, and hence has a least upper-bound, l, and $\{u_n\}$is bounded below, and has a greatest lower-bound, g. From part 1), we know that u_n and v_n converge to l and g, respectively. From $\lim u_n - v_n = 0$, it follows that $l = u$.

3. If not, then we can construct a subsequence of points $\{x_i\}$ such that either $\lim f(x_i) \leq 0$, as $x_i \to 0$. But then, $\lim_{x \to 0} f(x_i) \neq f(0)$, contradicting the continuity of f.

2 Linear Algebra

1. The rank of a matrix A is the dimension of the largest vector space spanned by the columns of A. For a set of vectors $\{v_i\}$ to span a vector space X, means that for every $x \in X$, there exists α_i such that $\sum \alpha_i v_i = x$.

2. Assume rank$(A) < m$. Then the reduced row-echelon form of A will have fewer than m pivot rows. We can construct a non-zero vector x in the kernel of A by setting to zero the indices of x that correspond to pivot rows, and set to 1 the entries of x that correspond to non-pivot rows. Now assume there is an $x \neq 0$, such that $\sum a_i x_i = 0$. This implies that a subset of the rows of A are linear combinations of other rows in A. As a result, the corresponding rows of the reduced row-echelon form of A will be zero, and the image of A in $\mathbb{R}^m$ will not include any vector with non-zero entries along those dimensions.

3 Inner product, norm

1. The forward direction is obvious: if $a = 0$, then $a^T x = 0$ for all x. Going the other way, if $a^T x = 0$ for all x, then, in particular, $a^T a = \sum a_i a_i = 0$, which is only true if $a_i = 0$ for all i.

   For the inquality, we'll first prove the contrapositive. Assume $a < 0$. Define $f(i) = 1$ if and only if $a_i < 0$. Then we can construct a vector x such that $x_i = 1$ if $f(i) = 1$, and $x_i = 0$ otherwise. Then $a^T x < 0$. That proves the backward direction. To prove the forward direction, assume there is an $x \geq 0$ such that $a^T x < 0$. Then $a_i < 0$ for some i, otherwise $a_i x_i \geq 0$ for all i, and $\sum a_i x_i$ would therefore be nonnegative. This proves the forward direction.

2. We substitute the expression for the Euclidean norm and simplify, to get:

$$\begin{aligned}
\frac{\|x+y\|^2 - \|x\|^2 - \|y\|^2}{2} &= \frac{\sum (x_i + y_i)(x_i + y_i) - \sum x_i x_i - \sum y_i y_i}{2} \\
&= \frac{\sum x_i x_i + \sum y_i y_i + 2 \sum x_i y_i - \sum x_i x_i - \sum y_i y_i}{2} \\
&= \sum x_i y_i \\
&= x^T y
\end{aligned}$$

   The proof of the other identity is essentially the same.

3. This is simply a matter of setting the identities equal and rearranging terms.

4 Multivariate Calculus

1. Let $u = x + \lambda d$. Then $f_{x,d}(\lambda) = f(x + \lambda d) = f(u(\lambda))$. Note that $u(\lambda)$ is a vector with elements that are funcntions of λ. Apply the chain rule to $f(u(\lambda))$ to get

$$\frac{df}{du}\frac{du}{d\lambda} = \nabla_u f \cdot d.$$

   Taking the derivative of $\nabla_u f \cdot d$ with respect to λ, we get

$$\left(\frac{d^2 f}{d^2 u}\frac{du}{d\lambda}\right) d = d^T H d,$$

   where H is the Hessian of f with respect to u.

2. We can write the ith element of ∇f as $\frac{\partial f}{\partial x_i}$. For $\|h\| < \epsilon$, we have that $f(x+h) \geq f(x)$. We can write the ith element of ∇f as

$$\frac{\partial f}{\partial x_i} = \lim_{h_i \to 0} \frac{f(x_i + h_i) - f(x_i)}{h_i}.$$

   For $h_i < 0$, $\frac{\partial f}{\partial x_i} \leq 0$, because the numerator is positive. For $h_i > 0, \frac{\partial f}{\partial x_i} \geq 0$. Because ∇f exists, each partial derivative exists, and so the limit must be the same for $h_i < 0$ and $h_i > 0$. It follows that $\frac{\partial f}{\partial x_i} = 0$, and so $\nabla f(x) = 0$.

3. Writing out the sum term-by-term, we get

$$f(x) = a^T x = \sum_i a_i x_i.$$

So the partial derivative with respect to x_i is

$$\frac{\partial f}{\partial x_i} = a_i,$$

so $\nabla f(x) = a$.

Doing the same for $x^T x$,we get

$$f(x) = x^T x = \sum_i x_i x_i.$$

So the partial derivative is

$$\frac{\partial f}{\partial x_i} = 2x_i.$$

So $\nabla f(x) = 2x$.

For $x^T M x$, the indices are a bit more involved, but the process is the same. We write

$$f(x) = x^T M x = \sum_j \sum_i x_i x_j m_{ij}.$$

Taking the partial with respect to x_j, we get

$$\frac{\partial f}{\partial x_j} = \sum_i x_i m_{ij}.$$

In matrix form, this is Mx.

Taking the partial with respect to x_i, we get

$$\frac{\partial f}{\partial x_i} = \sum_j x_j m_{ij}.$$

In matrix form, this is $M^T x$. Combining, we get

$$\nabla f(x) = \left(M + M^T\right) x.$$

5 Programming

See the file problem_5.py.