Recitation notes

April 4, 2008

1 Norms

1.1 l^p norms for $p \geq 1$ satisfy the triangle inequality (based on Royden)

We wish to show that $\|x+y\|_p \leq \|x\|_p + \|y\|_p$ for $p \geq 1$.

For $p = \infty$, this is $\max_i |x_i + y_i| \leq \max_i (|x_i| + |y_i|) \leq \max_i |x_i| + \max_i |y_i|$. $\square$

For $1 \leq p < \infty$, first assume without loss that $x, y \neq 0$ (the inequality is trivial if either vector is zero). Let $\alpha = \|x\|_p$, and $\beta = \|y\|_p$, from which we define $x_0 = \frac{1}{\alpha} x$, and $y_0 = \frac{1}{\beta} y$. Then:

$$\begin{aligned} \|x+y\|_p &= \|\alpha x_0 + \beta y_0\|_p \\ &= (\alpha+\beta) \left\| \frac{\alpha}{\alpha+\beta} x_0 + \frac{\beta}{\alpha+\beta} y_0 \right\|_p \\ &= (\alpha+\beta) \left(\sum_{k=1}^{d} \left| \frac{\alpha}{\alpha+\beta} x_{0,d} + \frac{\beta}{\alpha+\beta} y_{0,d} \right|^p \right)^{\frac{1}{p}} \end{aligned}$$

Since $f(x) = |x|^p$ is convex for $1 \leq p < \infty$, and noting that $\frac{\alpha}{\alpha+\beta}, \frac{\beta}{\alpha+\beta} \geq 0$ and $\frac{\alpha}{\alpha+\beta} + \frac{\beta}{\alpha+\beta} = 1$:

$$\begin{aligned} \|x+y\|_p &\leq (\alpha+\beta) \left(\sum_{k=1}^{d} \left(\frac{\alpha}{\alpha+\beta} |x_{0,d}|^p + \frac{\beta}{\alpha+\beta} |y_{0,d}|^p \right) \right)^{\frac{1}{p}} \\ &\leq (\alpha+\beta) \left(\frac{\alpha}{\alpha+\beta} \|x_0\|_p^p + \frac{\beta}{\alpha+\beta} \|y_0\|_p^p \right)^{\frac{1}{p}} \end{aligned}$$

Recalling that $\|x_0\|_p^p = \|y_0\|_p^p = 1$:

$$\|x+y\|_p \leq |\alpha+\beta|$$

WLOG But $\alpha = \|x\|_p$ and $\beta = \|y\|_p$, so:

$$\|x+y\|_p \leq \|x\|_p + \|y\|_p$$

$\square$

1.2 Norms are convex

A norm, by definition, satisfies the triangle inequality. Suppose that $\|\cdot\|$ is a norm. Then, if $\lambda \in [0,1]$:

$$\begin{aligned} \|\lambda x + (1-\lambda) y\| &\leq \|\lambda x\| + \|(1-\lambda) y\| \\ &\leq \lambda \|x\| + (1-\lambda) \|y\| \end{aligned}$$

2 Hessians

Suppose that $f : \mathbb{R}^d \to \mathbb{R}$ is twice differentiable. The *gradient* of f is the vector of partial derivatives:

$$\bigtriangledown f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_d} \end{bmatrix}$$

The *Hessian* is the matrix of partial second derivatives:

$$\bigtriangledown^2 f = \begin{bmatrix} \frac{\partial^2 f}{\partial^2 x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_d} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_d} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_d \partial x_1} & \frac{\partial^2 f}{\partial x_d \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_d^2} \end{bmatrix}$$

Note that since $\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$, this matrix is symmetric. It's important to realize that, like a derivative (or gradient), the Hessian is a *function* of the vector x–it takes on different values on different parts of the domain.

A word on notation: $\bigtriangledown^2$ is frequently used for the Laplacian operator $\bigtriangledown^2 f = \sum_{i=1}^{d} \frac{\partial^2 f}{\partial^2 x_i}$, but not in this class.

2.1 Positive definiteness

An eigenvalue is a scalar λ which satisfies the following:

$$Ax = \lambda x$$

For some $x \neq 0$ (which is called an eigenvector). Note that if the above holds for x, it holds for αx where $\alpha \neq 0$ is an arbitrary scalar. That is, the *scale* of eigenvectors is irrelevant. A $n \times n$ symmetric matrix will always have exactly n eigenvalue/eigenvector pairs, though the eigenvalues are not necessarily distinct. Additionally, the eigenvectors may always be taken to be orthogonal.

For small matrices, we may solve for the eigenvalues & eigenvectors explicitly:

$$\begin{aligned} \begin{bmatrix} 17 & -6 \\ -6 & 8 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} &= \lambda \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ \begin{bmatrix} 17x_1 - 6x_2 \\ -6x_1 + 8x_2 \end{bmatrix} &= \lambda \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ \begin{bmatrix} (17-\lambda) x_1 \\ (8-\lambda) x_2 \end{bmatrix} &= \begin{bmatrix} 6x_2 \\ 6x_1 \end{bmatrix} \end{aligned}$$

Soving for x_1 gives that $x_1 = \frac{6}{17-\lambda}x_2$ and $x_1 = \frac{8-\lambda}{6}x_2$, so that:

$$\begin{aligned}
\frac{6}{17-\lambda} &= \frac{8-\lambda}{6} \\
36 &= (8-\lambda)(17-\lambda) \\
0 &= 136 - 25\lambda + \lambda^2 - 36 \\
0 &= \lambda^2 - 25\lambda + 100 \\
\lambda &= \frac{25 \pm \sqrt{625-400}}{2} \\
&= \frac{25 \pm \sqrt{225}}{2} \\
&= \frac{25 \pm 15}{2} \\
&\in \{20, 5\}
\end{aligned}$$

A matrix is positive definite if all of its eigenvalues are strictly positive, and positive semi-definite if they are all nonnegative. Is the above matrix positive definite? (yes)

One simple test for positive (semi-)definiteness is that a matrix A is positive (semi-)definite if $x^T Ax$ is positive (nonnegative) for all $x \neq 0$.

A twice-differentiable function is convex if its Hessian is positive semidefinite everywhere, and is strictly convex if its Hessian is positive definite.

The definitions of negative (semi-) definite and (strictly) concave are the same, with the signs reversed.

2.2 Quadratic-over-linear is convex (page 73 of Boyd)

Consider $f(x,y) = \frac{x^2}{y}$ for $y > 0$. Then:

$$\begin{aligned}
\bigtriangledown f &= \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} \\
&= \begin{bmatrix} \frac{2x}{y} \\ -\frac{x^2}{y^2} \end{bmatrix}
\end{aligned}$$

And:

$$\begin{aligned}
\bigtriangledown^2 f &= \begin{bmatrix} \frac{\partial^2 f}{\partial^2 x} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial^2 y} \end{bmatrix} \\
&= \begin{bmatrix} \frac{2}{y} & -\frac{2x}{y^2} \\ -\frac{2x}{y^2} & \frac{2x^2}{y^3} \end{bmatrix}
\end{aligned}$$

Which we may write as:

$$\begin{aligned}
\bigtriangledown^2 f &= \frac{2}{y^3} \begin{bmatrix} y^2 & -xy \\ -xy & x^2 \end{bmatrix} \\
&= \frac{2}{y^3} \begin{bmatrix} y \\ -x \end{bmatrix} \begin{bmatrix} y & -x \end{bmatrix}
\end{aligned}$$

By inspection, one eigenvector of $\begin{bmatrix} y^2 & -xy \\ -xy & x^2 \end{bmatrix}$ is $\begin{bmatrix} y \\ -x \end{bmatrix}$, with eigenvalue $y^2 + x^2 \geq 0$:

$$\begin{aligned} \begin{bmatrix} y^2 & -xy \\ -xy & x^2 \end{bmatrix} \begin{bmatrix} y \\ -x \end{bmatrix} &= \begin{bmatrix} y \\ -x \end{bmatrix} \begin{bmatrix} y & -x \end{bmatrix} \begin{bmatrix} y \\ -x \end{bmatrix} \\ &= \begin{bmatrix} y \\ -x \end{bmatrix} \left(y^2 + x^2\right) \end{aligned}$$

While the other eigenvalue is zero, since $\begin{bmatrix} y^2 & -xy \\ -xy & x^2 \end{bmatrix}$ has rank 1 (eigenvector $\begin{bmatrix} x \\ y \end{bmatrix}$). Since $\frac{2}{y^3} \geq 0$ for $y > 0$, this gives that:

$$\nabla^2 f \succeq 0$$

Showing that f is convex. Is it strictly convex? (no)

2.3 Arithmetic-geometric means inequality

Consider $f(x) = \left(\prod_{i=1}^n x_i\right)^{\frac{1}{n}}$ for $x > 0$. We may write $\log f(x) = \frac{1}{n} \sum_{i=1}^n \log x_i$. By the concavity of the logarithm:

$$\frac{1}{n} \sum_{i=1}^n \log x_i \leq \log \left(\frac{1}{n} \sum_{i=1}^n x_i \right)$$

Exponentiating both sides gives:

$$f(x) \leq \frac{1}{n} \sum_{i=1}^n x_i$$

Which is the arithmetic-geometric means inequality.

2.4 Geometric mean is concave (page 74 of Boyd)

Again consider $f(x) = \left(\prod_{i=1}^n x_i\right)^{\frac{1}{n}}$ for $x > 0$. Its gradient is:

$$\begin{aligned} \frac{\partial f}{\partial x_k} &= \frac{1}{n} \left(\prod_{i=1}^n x_i \right)^{\frac{1}{n} - 1} \frac{\prod_{i=1}^n x_i}{x_k} \\ &= \frac{1}{n x_k} \left(\prod_{i=1}^n x_i \right)^{\frac{1}{n}} \end{aligned}$$

And its Hessian is:

$$\begin{aligned}
\frac{\partial^2 f}{\partial^2 x_k} &= \left(\frac{1}{nx_k}\right)^2 \left(\prod_{i=1}^n x_i\right)^{\frac{1}{n}} - \frac{1}{nx_k^2}\left(\prod_{i=1}^n x_i\right)^{\frac{1}{n}} \\
&= \left(\frac{1}{n^2} - \frac{1}{n}\right)\frac{1}{x_k^2}\left(\prod_{i=1}^n x_i\right)^{\frac{1}{n}} \\
&= (1-n)\frac{1}{n^2 x_k^2}\left(\prod_{i=1}^n x_i\right)^{\frac{1}{n}} \\
\frac{\partial^2 f}{\partial x_k \partial x_j} &= \frac{1}{n^2 x_k x_j}\left(\prod_{i=1}^n x_i\right)^{\frac{1}{n}}
\end{aligned}$$

Which we may write as:

$$\begin{aligned}
\bigtriangledown^2 f &= -\frac{1}{n^2}\left(\prod_{i=1}^n x_i\right)^{\frac{1}{n}} \begin{bmatrix} \frac{n-1}{x_1^2} & -\frac{1}{x_1 x_2} & \cdots & -\frac{1}{x_1 x_n} \\ -\frac{1}{x_2 x_1} & \frac{n-1}{x_2^2} & \cdots & -\frac{1}{x_2 x_n} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{x_n x_1} & -\frac{1}{x_n x_2} & \cdots & \frac{n-1}{x_n^2} \end{bmatrix} \\
&= -\frac{1}{n^2}\left(\prod_{i=1}^n x_i\right)^{\frac{1}{n}} \left(\begin{bmatrix} \frac{n}{x_1^2} & 0 & \cdots & 0 \\ 0 & \frac{n-1}{x_2^2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{n-1}{x_n^2} \end{bmatrix} - \begin{bmatrix} \frac{1}{x_1^2} & \frac{1}{x_1 x_2} & \cdots & \frac{1}{x_1 x_n} \\ \frac{1}{x_2 x_1} & \frac{1}{x_2^2} & \cdots & \frac{1}{x_2 x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{x_n x_1} & \frac{1}{x_n x_2} & \cdots & \frac{1}{x_n^2} \end{bmatrix} \right) \\
&= -\frac{1}{n^2}\left(\prod_{i=1}^n x_i\right)^{\frac{1}{n}} \left(n\mathrm{diag}\left(\frac{1}{x_1^2}, \frac{1}{x_2^2}, \ldots, \frac{1}{x_n^2}\right) - \begin{bmatrix} \frac{1}{x_1} \\ \frac{1}{x_2} \\ \vdots \\ \frac{1}{x_n} \end{bmatrix} \begin{bmatrix} \frac{1}{x_1} \\ \frac{1}{x_2} \\ \vdots \\ \frac{1}{x_n} \end{bmatrix}^T \right)
\end{aligned}$$

Consider $v^T \bigtriangledown^2 f v$:

$$v^T \left(\bigtriangledown^2 f\right) v = -\frac{1}{n^2}\left(\prod_{i=1}^n x_i\right)^{\frac{1}{n}} \left(n \sum_{i=1}^n \frac{v_i^2}{x_i^2} - \left(\sum_{i=1}^n \frac{v_i}{x_i}\right)^2 \right)$$

Note that we may write $\sum_{i=1}^n \frac{v_i^2}{x_i^2}$ as:

$$\sum_{i=1}^n \frac{v_i^2}{x_i^2} = \left\| \begin{bmatrix} \frac{v_1}{x_1} \\ \frac{v_2}{x_2} \\ \vdots \\ \frac{v_n}{x_n} \end{bmatrix} \right\|_2^2$$

And by the Cauchy-Schwarz inequality ($\left(a^T b\right)^2 \leq \|a\|_2^2 \|b\|_2^2$) with $a = \mathbf{1}$ and $b_i = \frac{v_i}{x_i}$:

$$\left(\sum_{i=1}^n \frac{v_i}{x_i}\right)^2 \leq n \sum_{i=1}^n \frac{v_i^2}{x_i^2}$$

So that $v^T \left(\bigtriangledown^2 f\right) v \leq 0$, showing that f is concave.