Gradients of Inner Products

Let

\[ \phi(x) = c^T x \]

where \( c \) is a constant vector. Then, from

\[ \phi(x) = \sum_{j=1}^{n} c_j x_j, \]

we have

\[ \frac{\partial \phi}{\partial x_k} = \sum_{j=1}^{n} c_j \frac{\partial x_j}{\partial x_k} = \sum_{j=1}^{n} c_j \delta_{jk} = c_k, \]

and therefore

\[ \nabla \phi(x) = c. \]

Now, let

\[ \varphi(x) = x^T B x = \sum_{i=1}^{n} \sum_{j=1}^{n} b_{ij} x_i x_j. \]

Then

\[ \frac{\partial \varphi}{\partial x_k} = \sum_{i=1}^{n} \sum_{j=1}^{n} b_{ij} \frac{\partial (x_i x_j)}{\partial x_k} \]

\[ = \sum_{i=1}^{n} \sum_{j=1}^{n} b_{ij} (\delta_{ik} x_j + x_i \delta_{jk}) \]

\[ = \sum_{i=1}^{n} \sum_{j=1}^{n} b_{ij} x_j \delta_{ik} + \sum_{i=1}^{n} \sum_{j=1}^{n} b_{ij} x_i \delta_{jk} \]

\[ = \sum_{j=1}^{n} b_{kj} x_j + \sum_{i=1}^{n} b_{ik} x_i \]

\[ = (Bx)_k + \sum_{i=1}^{n} (B^T)_{ki} x_i \]

\[ = (Bx)_k + (B^T x)_k. \]

We conclude that

\[ \nabla \varphi(x) = (B + B^T)x. \]
Gradient of the 2-Norm of the Residual Vector

From
\[ \|x\|_2 = \sqrt{x^T x}, \]
and the properties of the transpose, we obtain
\[
\|b - Ax\|_2^2 = (b - Ax)^T (b - Ax) = b^T b - (Ax)^T b - b^T Ax + x^T A^T Ax
\]
\[
= b^T b - 2b^T Ax + x^T A^T Ax
\]
Using the formulas from the previous section, with \( c = A^T b \) and \( B = A^T A \), we have
\[
\nabla(\|b - Ax\|_2^2) = -2A^T b + (A^T A)^T x.
\]
However, because
\[
(A^T A)^T = A^T (A^T)^T = A^T A,
\]
this simplifies to
\[
\nabla(\|b - Ax\|_2^2) = -2A^T b + 2A^T Ax = 2(A^T Ax - A^T b).
\]

Positive Definiteness of \( A^T A \)

Let \( A \) be \( m \times n \), with \( m \geq n \) and \( \text{rank}(A) = n \). Then, if \( x \neq 0 \), it follows from the linear independence of \( A \)'s columns that \( Ax \neq 0 \). We then have
\[
x^T A^T Ax = (Ax)^T Ax = \|Ax\|_2^2 > 0,
\]
since the norm of a nonzero vector must be positive. It follows that \( A^T A \) is not only symmetric, but positive definite as well.

Hessians of Inner Products

The Hessian of the function \( \varphi(x) \), denoted by \( H_{\varphi}(x) \), is the matrix with entries
\[
h_{ij} = \frac{\partial^2 \varphi}{\partial x_i \partial x_j}.
\]
Because mixed second partial derivatives satisfy
\[
\frac{\partial^2 \varphi}{\partial x_i \partial x_j} = \frac{\partial^2 \varphi}{\partial x_j \partial x_i}
\]
as long as they are continuous, the Hessian is symmetric under these assumptions.

In the case of \( \varphi(x) = x^T B x \), whose gradient is \( \nabla \varphi(x) = (B + B^T)x \), the Hessian is \( H_{\varphi}(x) = B + B^T \). It follows from the previously computed gradient of \( \|b - Ax\|_2^2 \) that its Hessian is \( 2A^T A \). Therefore, the Hessian is positive definite, which means that the unique critical point \( x \), the solution to the normal equations \( A^T Ax - A^T b = 0 \), is a minimum.

In general, if the Hessian at a critical point is
• positive definite, meaning that its eigenvalues are all positive, then the critical point is a local minimum.
• negative definite, meaning that its eigenvalues are all negative, then the critical point is a local maximum.
• indefinite, meaning that it has both positive and negative eigenvalues, then the critical point is a saddle point.
• singular, meaning that one of its eigenvalues is zero, then the second derivative test is inconclusive.

The Condition Number of $A^T A$

When $A$ is $n \times n$ and invertible,

$$\kappa_2(A) = \|A\|_2 \|A^{-1}\|_2 = \frac{\sigma_1}{\sigma_n},$$

where $\sigma_1$ and $\sigma_n$ are the largest and smallest singular values, respectively, of $A$. If $A$ is $m \times n$ with $m > n$ and rank($A$) = $n$, $A^{-1}$ does not exist, but the quantity $\sigma_1/\sigma_n$ is still defined and an appropriate measure of the sensitivity of the least squares problem to perturbations in the data, so we denote this ratio by $\kappa_2(A)$ in this case as well.

From the relations

$$Av_j = \sigma_j u_j, \quad A^T u_j = \sigma_j v_j, \quad j = 1, \ldots, n,$$

where $u_j$ and $v_j$ are the left and right singular vectors, respectively, of $A$, we have

$$A^T Av_j = \sigma_j^2 v_j.$$

That is, $\sigma_j^2$ is an eigenvalue of $A^T A$. Furthermore, because $A^T A$ is symmetric positive definite, the eigenvalues of $A^T A$ are also its singular values. Specifically, if $A = U \Sigma V^T$ is the SVD of $A$, then $V^T (\Sigma^T \Sigma)V$ is the SVD of $A^T A$.

It follows that the condition number in the 2-norm of $A^T A$ is

$$\kappa_2(A^T A) = \|A^T A\|_2 \|A^T A^{-1}\|_2 = \frac{\sigma_1^2}{\sigma_n^2} = \left(\frac{\sigma_1}{\sigma_n}\right)^2 = \kappa_2(A)^2.$$

Note that because $A$ has full column rank, $A^T A$ is nonsingular, and therefore $(A^T A)^{-1}$ exists, even though $A^{-1}$ may not.