The Hessian of a Cost Function

Herman Jaramillo

May 10, 2016

1 The problem

Given the least square cost function

$$\mathcal{C}(p) = \frac{1}{2}\langle u(p) - d\ ,\ u(p) - d\rangle$$

This is pure algorthmic. we do not define the Hessian rigurously.

Pratt et. al., [1] show a compact representation of FWI problem using matrices and in their representation they find ways to compute the Hessian.

Here we formulate the problem in terms of functionals and provide new insights into details about the computations.

The Hessian is the derivative of the gradient. The gradient is

$$\nabla_p(\mathcal{C}(p)) = \langle u_p\ ,\ u - d\rangle$$

assuming that we are working over the real spaces.

Then

$$(\nabla_p \mathcal{C}(p))_p = \langle u_p\ ,\ u_p\rangle + \langle (u_p)_p\ ,\ u - d\rangle.$$

Pratt et. al., call $u_d = J$ where J is for Jacobian. In matrix notation this is

$$H = (\nabla_p \mathcal{C}(p))_p = J^T J + (dJ/dp)^T(u - d) = H_1 + H_2.$$

The first term (H_1) is used as an approximation of the Gauss–Newton interation technique, while both terms (H) provide the full Newton inversion. If $u = Ap$, then $J = A$ and $\nabla_p(\mathcal{C}(p))_p = A^T A$. This is the reason why sometimes abuse the language saying that A^TA is the Hessian, for the least square approximation of $u \approx Ap$.

Let us assume we know u_p. Then H_1 is easy to compute as

$$H_1 = J^TJ.$$

Now, for H_2, tarting at the modeling equation:

$$A(p)u = f.$$

Take first derivative

$$A_pu + Au_p = 0,$$

and second derivative

$$(A_p)_pu + A_pu_p + A_pu_p + A(u_p)_p = 0,$$

so

$$A(u_p)_p = -(A_p)_pu - 2A_pu_p$$

As in the gradient computation, let us choose v such that

$$A^*v = u - d,$$

(the same already obtained by the gradient method) so

$$\langle A(u_p)_p\ ,\ v\rangle = \langle (u_p)_p\ ,\ u - d\rangle = H_2,$$

and the second part of the Hessian is computed with the equation

$$\begin{aligned} H_2 &= \langle -(A_p)_pu - 2A_pu_p\ ,\ v\rangle \\ &= -\langle\ (A_p)_pu,\ v\rangle - 2\langle\ A_pu_p,\ v\rangle \\ &= -\langle\ u,\ [(A_p)_p]^*v\rangle - 2\langle\ u_p,\ [A_p]^*v\rangle \end{aligned}$$

But, again, this needs that we know u_p and the computation of u_p is very expensive.

References

[1] Pratt R. G., Shin G., and Hicks G. J. Gauss–newton and full newton methods in frequency–space seismic waveform inversion. *Geophys. J. Int.*, (133):341–362, 1998.