Newton-Raphson step. Quadratic expansion

$$E \approx E_0 + \Delta\mathbf{x}^{\mathrm{t}}\mathbf{g}_0 + \tfrac{1}{2}\Delta\mathbf{x}^{\mathrm{t}}\mathbf{H}_0\Delta\mathbf{x} \qquad \Delta\mathbf{x} \equiv \mathbf{x} - \mathbf{x}_0$$

Linear gradient expansion

$$\mathbf{g} = \mathbf{g}_0 + \mathbf{H}\Delta\mathbf{x}_0$$

Setting the gradient at the next point to zero gives the Newton-Raphson step

$$\mathbf{g} \overset{!}{=} \mathbf{0} \implies \Delta\mathbf{x} = -\mathbf{H}_0^{-1}\mathbf{g}_0$$

For a perfectly quadratic surface, the Hessian is constant and the Newton-Raphson step takes us directly to the stationary point. On a surface with cubic and higher-order terms, this step can be repeated iteratively until we are close enough to the stationary point that the region separating us from it is approximately quadratic.

The secant condition. For points $\mathbf{x}$ and $\mathbf{x}_0$ sharing a locally quadratic region with each other, the change in the gradient between them is described by the following.

$$\mathbf{H}\Delta\mathbf{x} \approx \mathbf{H}_0\Delta\mathbf{x} \approx \Delta\mathbf{g} \qquad \Delta\mathbf{g} \equiv \mathbf{g} - \mathbf{g}_0$$

This can be used to determine an approximation $\tilde{\mathbf{H}}$ to the Hessian at $\mathbf{x}$ using the Hessian at $\mathbf{x}_0$. Namely, we require the approximation to satisfy

$$\tilde{\mathbf{H}}\Delta\mathbf{x} \overset{!}{=} \Delta\mathbf{g} \qquad \tilde{\mathbf{H}} = \mathbf{H}_0 + \Delta\tilde{\mathbf{H}}$$

which is known as the *quasi-Newton condition*, or alternatively the *secant condition*. If the dimension is d, then we have d linear equations and d^2 undetermined matrix elements (or rather $d(d+1)/2$ matrix elements, by symmetry), so the system is underdetermined – we have an infinite number of solutions.

The simplest way forward is to assume $\Delta\tilde{\mathbf{H}}$ has rank 1, which for a symmetric matrix implies that it has the form

$$\Delta\tilde{\mathbf{H}} = \eta\,\mathbf{e}\mathbf{e}^{\mathrm{t}}$$

where η is a scalar and $\mathbf{e}$ is a unit vector. This is the *symmetric rank-1* (SR1) approximation. Substituting this into the secant condition, we find

$$\mathbf{e} = \frac{\Delta\mathbf{g} - \mathbf{H}_0\Delta\mathbf{x}}{\|\Delta\mathbf{g} - \mathbf{H}_0\Delta\mathbf{x}\|} \qquad \eta = \frac{\mathbf{e}\cdot(\Delta\mathbf{g} - \mathbf{H}_0\Delta\mathbf{x})}{\mathbf{e}\cdot\Delta\mathbf{x}} = \frac{(\Delta\mathbf{g} - \mathbf{H}_0\Delta\mathbf{x})^2}{(\Delta\mathbf{g} - \mathbf{H}_0\Delta\mathbf{x})\cdot\Delta\mathbf{x}}$$

which allows us to simplify the SR1 matrix update.

$$\Delta\tilde{\mathbf{H}}_{\mathrm{SR1}} \equiv \frac{(\Delta\mathbf{g} - \mathbf{H}_0\Delta\mathbf{x})(\Delta\mathbf{g} - \mathbf{H}_0\Delta\mathbf{x})^{\mathrm{t}}}{(\Delta\mathbf{g} - \mathbf{H}_0\Delta\mathbf{x})\cdot\Delta\mathbf{x}}$$

This provides a scheme for updating the Hessian using the change in the first derivative as we move across the potential surface. If we take several steps in a small, locally quadratic region of the surface, this approximation will improve with each step, and eventually we will have quite a good approximation to $\mathbf{H}$. We can determine other approximations by assuming that the Hessian correction has rank 2, which can be used to derive the *Broyden-Fletcher-Goldfarb-Shanno* (BFGS) and *Davidon-Fletcher-Powell* (DFP) Hessian updates.

Combining secant updating with Newton-Raphson optimization leads to the *quasi-Newton optimization algorithms*, which have the following form.

1. Evaluate the gradient $\mathbf{g}_0$ at the starting point $\mathbf{x}_0$ and either evaluate the Hessian or use an approximation. Setting $\tilde{\mathbf{H}}_0 = \frac{\|\mathbf{g}_0\|}{s_0}\mathbf{1}$ will yield a gradient descent of length s_0 for the initial step.

2. Take a Newton-Raphson step.

$$\mathbf{x} = \mathbf{x}_0 + \Delta\mathbf{x} \qquad \Delta\mathbf{x} = -\mathbf{H}_0^{-1}\mathbf{g}_0$$

3. Evaluate the gradient $\mathbf{g}$ at the new point. If $\max(\mathbf{g}) < \text{tol}$, quit. We have converged.

4. Otherwise, update the Hessian using a secant method like SR1.

$$\tilde{\mathbf{H}} = \tilde{\mathbf{H}}_0 + \Delta\tilde{\mathbf{H}}_{\text{SR1}} \qquad \Delta\tilde{\mathbf{H}}_{\text{SR1}} = \frac{(\Delta\mathbf{g} - \mathbf{H}_0\Delta\mathbf{x})(\Delta\mathbf{g} - \mathbf{H}_0\Delta\mathbf{x})^{\text{t}}}{(\Delta\mathbf{g} - \mathbf{H}_0\Delta\mathbf{x}) \cdot \Delta\mathbf{x}}$$

5. Set $\mathbf{x}_0 \leftarrow \mathbf{x}, \mathbf{g}_0 \leftarrow \mathbf{g}$, and $\mathbf{H}_0 \leftarrow \mathbf{H}$ and return to step 2.