Nonlinear solutions (following Rodgers Ch.5)

We have derived the linear Gaussian variational approach equation

$$\hat{x} = x_a + \left(K^T S_\varepsilon^{-1} K + S_a^{-1}\right)^{-1} K^T S_\varepsilon^{-1}\left(y - Kx_a\right) \tag{1}$$

We can generalize this to include some range of nonlinear behavior, so-called moderately nonlinear behavior, by substituting $F(x)$ for Kx. So the linear form

$$-2\ln P(x|y) = (y - Kx)^T S_\varepsilon^{-1}(y - Kx) + (x - x_a)^T S_a^{-1}(x - x_a) + c_3 \tag{2}$$

becomes the nonlinear form

$$-2\ln P(x|y) = \left[y - F(x)\right]^T S_\varepsilon^{-1}\left[y - F(x)\right] + \left[x - x_a\right]^T S_a^{-1}\left[x - x_a\right] + c_3 \tag{3}$$

The solution to this equation that we want is the state vector, x, that maximizes the probability in (3). The approach is to equate the derivative of (3) to zero and find the solution

$$\nabla_x\left\{-2\ln P(x|y)\right\} = -\left[\nabla_x F(x)\right]^T S_\varepsilon^{-1}\left[y - F(x)\right] + S_a^{-1}\left[x - x_a\right] = 0 \tag{4}$$

We write $K(x) = \nabla_x F(x)$ which is still dy/dx. Now because of the nonlinear behavior, K is evaluated at x and therefore depends on x which in the linear form did not. So the optimum or maximum probability state, $\hat{x}$, is the solution to the equation

$$-\hat{K}^T(\hat{x}) S_\varepsilon^{-1}\left[y - F(\hat{x})\right] + S_a^{-1}\left[\hat{x} - x_a\right] = 0 \tag{5}$$

This must be solved iteratively.

If the problem is not too nonlinear, one can use Newtonian iteration to find the zero of a gradient of the cost function, (3). For the general vector equation, $g(x)=0$, the iteration, along the lines of Newton's method is

$$x_{i+1} = x_i - \left[\nabla_x g(x_i)\right]^{-1} g(x_i) \tag{6}$$

Taking (5) to be $g(x)=0$, the gradient of g is

$$\nabla_x g = S_a^{-1} + K^T S_\varepsilon^{-1} K - \left[\nabla_x K^T\right] S_\varepsilon^{-1}\left[y - F(x)\right] \tag{7}$$

Since $g(x)$ is the derivative of the cost function (3), $\nabla_x g$ is the second derivative of the cost function called the *Hessian*. The Hessian is expensive to calculate because it not only includes the Jacobean K which is the first derivative of the forward model but it also includes the second derivative of the forward model which can be very complicated and expensive to evaluate.

The right hand term in (7) is this expensive term and in practice is dropped which is OK as long as the situation is not too nonlinear. Under these conditions, the approximation is made

$$\nabla_x g \cong S_a^{-1} + K^T S_\varepsilon^{-1} K \tag{8}$$

Substituting (5) and (8) into (6) we obtain

$$x_{i+1} = x_i + \left(S_a^{-1} + K_i^T S_\varepsilon^{-1} K_i\right)^{-1}\left[K_i^T S_\varepsilon^{-1}\left(y - F(x_i)\right) - S_a^{-1}\left(x_i - x_a\right)\right] \tag{9}$$

which can be rewritten as a departure from x_a rather than x_i as follows

$$x_{i+1} = x_a + \left(S_a^{-1} + K_i^T S_\varepsilon^{-1} K_i\right)^{-1} K_i^T S_\varepsilon^{-1}\left[\left(y - F\left(x_i\right)\right) + K_i\left(x_i - x_a\right)\right] \tag{10}$$

where it is convenient to start the iterations with $x_i = x_a$. The first iteration in this case is

$$x_{i+1} = x_a + \left(S_a^{-1} + K_a^T S_\varepsilon^{-1} K_a\right)^{-1} K_a^T S_\varepsilon^{-1}\left[\left(y - F\left(x_a\right)\right)\right] \tag{11}$$

which looks very much like (1).