Derivation of FORM algorithm, by Dr. Chris Hoyle, April 2016.

The problem to be solved is:

$$\begin{aligned}&\textit{Minimize }\ \mathbf{u}^{\mathrm{T}}\mathbf{u}\\&\textit{Subject to }\ G(\mathbf{u})=0\end{aligned}$$

We cannot solve it analytically, so we are using a **Numerical Method** for Optimization. Note that since this is a numerical method, we are *iteratively* searching for the value of **u** to minimize $\mathbf{u}^{\mathrm{T}}\mathbf{u}$ subject to the constraint of staying on the **G(u)** = 0 line. We do this by starting at some initial point **u** in iteration *k*, and then searching in some direction $\Delta\mathbf{u}$ in the next iteration, *k*+1, until we reach some *convergence criterion*. The updating of **u** from one iteration to the next is given by:

$$\mathbf{u}^{k+1}=\mathbf{u}^{k}+\Delta\mathbf{u}$$

The numerical method for optimization we will use is the **Constrained Steepest Descent (CSD)** Algorithm (Arora, 2012). The form shown here is slightly modified to make the math cleaner, but the modification is mathematically equivalent to the CSD in Arora:

Minimize: $f^{k}+\nabla f^{k,T}d+d^{T}d$

Subject to: $\nabla G\left(\mathbf{u}^{k}\right)^{T}d+G\left(\mathbf{u}^{k}\right)=0$

In this equation, *f* is the objective function which $\mathbf{u}^{\mathrm{T}}\mathbf{u}$, *d* is the search direction which is $\Delta\mathbf{u}$, and *k* is the iteration number. Note that T indicates **Vector Transpose** and we use the convention that all vectors are column vectors (and their transpose is a row vector).

We can rewrite the objective function $\mathbf{u}^{\mathrm{T}}\mathbf{u}$ using the formulation above ($f^{k}+\nabla f^{k,T}d+d^{T}d$) making use of the fact that the derivative of $\mathbf{u}^{\mathrm{T}}\mathbf{u}$ = 2**u,** we get**:**

$$\text{Minimize: } \mathbf{u}^{\mathrm{T}}\mathbf{u}+2\mathbf{u}^{\mathrm{T}}\Delta\mathbf{u}+\Delta\mathbf{u}^{\mathrm{T}}\Delta\mathbf{u}$$

Recognizing by vector algebra that $\mathbf{u}^{\mathrm{T}}\Delta\mathbf{u}=\Delta\mathbf{u}^{\mathrm{T}}\mathbf{u}$**,** we get the following for our objective (and adding the index *k*):

$$\text{Minimize: } \mathbf{u}^{\mathrm{T}}\mathbf{u}+2\mathbf{u}^{\mathrm{T}}\Delta\mathbf{u}+\Delta\mathbf{u}^{\mathrm{T}}\Delta\mathbf{u}=(\mathbf{u}^{\mathrm{T}}+\Delta\mathbf{u}^{\mathrm{T}})\,(\mathbf{u}+\Delta\mathbf{u})=(\mathbf{u}^{k}+\Delta\mathbf{u})^{\mathrm{T}}(\mathbf{u}^{k}+\Delta\mathbf{u})$$

We can now write the *Lagrangian* function for CSD problem as follows (combined objective and constraint):

$$L=(\mathbf{u}^{\boldsymbol{k}}+\Delta\mathbf{u})^{T}(\mathbf{u}^{\boldsymbol{k}}+\Delta\mathbf{u})\ +\lambda\left(\nabla G\left(\mathbf{u}^{k}\right)^{T}\Delta\mathbf{u}+G\left(\mathbf{u}^{k}\right)\right)$$

In this equation, λ is the *Lagrange multiplier*. We have 2 unknowns, λ and $\Delta\mathbf{u}$**.** We need to find the *stationary points* of the Lagrangian to solve our problem, so we take a partial derivative of *L* wrt λ and $\Delta\mathbf{u}$ to get 2 equations with 2 unknowns**.** We use our previous matrix rule that $\mathbf{u}^{\mathrm{T}}\mathbf{u}$ = 2**u** to get**:**

$$\frac{\partial L}{\partial\Delta\mathbf{u}}=2\left(\mathbf{u}^{\boldsymbol{k}}+\Delta\mathbf{u}\right)+\lambda\nabla G\left(\mathbf{u}^{k}\right)=0 \qquad (1)$$

$$\frac{\partial L}{\partial\lambda}=\nabla G\left(\mathbf{u}^{k}\right)^{T}\Delta\mathbf{u}+G\left(\mathbf{u}^{k}\right)=0 \qquad (2)$$

What we are going to do now is solve for λ in equation 1. Start by rearranging the equation:

Derivation of FORM algorithm, by Dr. Chris Hoyle, April 2016.

$$-2\left(\mathbf{u}^{\boldsymbol{k}}+\Delta\mathbf{u}\right)=\lambda\nabla G\left(\mathbf{u}^{k}\right)$$

We've run into an issue! To get rid of $\nabla G\left(\mathbf{u}^{k}\right)$ from the right side and solve for λ, *we can't simply divide* by $\nabla G\left(\mathbf{u}^{k}\right)$. The reason is that since $\nabla G\left(\mathbf{u}^{k}\right)$ is a vector, we are technically taking the **inverse** of the vector, i.e. $\nabla G\left(\mathbf{u}^{k}\right)^{-1}$, which doesn't exist. We can only invert certain square matrices, not vectors. So we need to turn $\nabla G\left(\mathbf{u}^{k}\right)$ *either* into a square matrix or a scalar somehow. Luckily, we can turn it into a scalar by multiplying it by $\nabla G\left(\mathbf{u}^{k}\right)^{T}$, so we will multiply both sides $\nabla G\left(\mathbf{u}^{k}\right)^{T}$, and then divide through both sides by $\nabla G\left(\mathbf{u}^{k}\right)^{T}\nabla G\left(\mathbf{u}^{k}\right)$, which is a scalar quantity (i.e. row vector times a column vector). We get the following:

$$-\frac{2}{\nabla G\left(\mathbf{u}^{k}\right)^{T}\nabla G\left(\mathbf{u}^{k}\right)}\nabla G\left(\mathbf{u}^{k}\right)^{T}\left(\mathbf{u}^{\boldsymbol{k}}+\Delta\mathbf{u}\right)=\lambda \tag{3}$$

We've now solved for λ! One issue remaining: we still have $\Delta\mathbf{u}$ in the equation, which is an unknown. We can use equation 2 above to get rid of $\Delta\mathbf{u}$ in this equation**.** Rearranging equation 2, we get:

$$\nabla G\left(\mathbf{u}^{k}\right)^{T}\Delta\mathbf{u}=-G\left(\mathbf{u}^{k}\right) \tag{4}$$

We can rewrite Equation 3 for λ as:

$$\lambda=-\frac{2}{\nabla G(\mathbf{u}^{k})^{T}\nabla G(\mathbf{u}^{k})}\nabla G\left(\mathbf{u}^{k}\right)^{T}\mathbf{u}^{\boldsymbol{k}}+\nabla G\left(\mathbf{u}^{k}\right)^{T}\Delta\mathbf{u}$$

and then substitute in Equation 4:

$$\lambda=-\frac{2}{\nabla G(\mathbf{u}^{k})^{T}\nabla G(\mathbf{u}^{k})}\nabla G\left(\mathbf{u}^{k}\right)^{T}\mathbf{u}^{\boldsymbol{k}}-G\left(\mathbf{u}^{k}\right)$$

Now we substitute this expression for λ back into equation 1:

$$2\left(\mathbf{u}^{\boldsymbol{k}}+\Delta\mathbf{u}\right)+\left[-\frac{2}{\nabla G(\mathbf{u}^{k})^{T}\nabla G(\mathbf{u}^{k})}\nabla G\left(\mathbf{u}^{k}\right)^{T}\mathbf{u}^{\boldsymbol{k}}-G\left(\mathbf{u}^{k}\right)\right]\nabla G\left(\mathbf{u}^{k}\right)=0$$

Now we can solve easily for the quantity of interest $\left(\mathbf{u}^{\boldsymbol{k}}+\Delta\mathbf{u}\right)$, which is in fact equal to $\mathbf{u}^{k+1}$, exactly what we need for the next point in our algorithm!:

$$\left(\mathbf{u}^{\boldsymbol{k}}+\Delta\mathbf{u}\right)=\left[\frac{1}{\nabla G(\mathbf{u}^{k})^{T}\nabla G(\mathbf{u}^{k})}\nabla G\left(\mathbf{u}^{k}\right)^{T}\mathbf{u}^{\boldsymbol{k}}-G\left(\mathbf{u}^{k}\right)\right]\nabla G\left(\mathbf{u}^{k}\right)$$

Or:

$$\mathbf{u}^{k+1}=\left[\frac{\left\{\nabla G\left(\mathbf{u}^{k}\right)^{T}\mathbf{u}^{\boldsymbol{k}}-G\left(\mathbf{u}^{k}\right)\right\}}{\nabla G(\mathbf{u}^{k})^{T}\nabla G(\mathbf{u}^{k})}\right]\nabla G\left(\mathbf{u}^{k}\right)$$

All done!