Mixed FEM stuff

1 Problem Setup

1.1 Continuous Problem

We begin with our continuous mixed energy. We write the energy as an integral over a reference domain $\Omega \subset \mathbb{R}^d$. For notation we denote vectors by bold lower case letters, matrices by upper-case letters, and scalars by lower case letters. Our mixed energy is a function of deformed coordinates, $\mathbf{x} \in \mathbb{R}^d$, symmetric deformation matrices $S \in \mathbb{R}^{d \times d}$, and Lagrange multipliers $\Lambda \in \mathbb{R}^{d \times d}$. Our continuous mixed variational is. **(Note: this is a funny semi-discretization - we've discretized in time so far – using Backwards Euler but not space - notice a lo of FE papers do the opposite to get an ODE first by discretizing in space first. No reason we need to start with this vs the other - right? Also we could propose the setup in the fully continuous setting first too - right?)**

$$\mathcal{L}(\mathbf{x}, S, \Lambda) = \int_\Omega \frac{\rho}{2h^2}||\mathbf{x} - \tilde{\mathbf{x}}||_2^2 + \Psi(S) - \Lambda : (R(\mathbf{x})S - F(\mathbf{x}))dV \tag{1}$$

where h is the timestep, ρ is the density, and $\tilde{\mathbf{x}} = \mathbf{x}^t + h\dot{\mathbf{x}}^t + h^2\mathbf{f}_{\text{ext}}$. In the mixed formulation, our strain energy density function Ψ is now a function of S rather than the deformed coordinates. R is the rotation matrix from the polar decomposition of the deformation gradient, $F(\mathbf{x}) = R(\mathbf{x})S(\mathbf{x})$. **Where specifically we have $R(F(x))$ to indicate R is a projection of the deformation gradient onto SO(3)).**

1.2 Discretized Energy

We represent the domain by $\mathcal{T} \equiv \{K_i\}_{i=1}^{n_e}$ consisting of a set of elements $K_1, \ldots, K_{n_e}$. **(Minor - me being fussy - but can we use lower case for indexing?)** Per-element variables are represent by a subscript K. Our deformed coordinates are discretized on a nodal mesh, such as triangles or tetrahedra, and are stored in the vector $\mathbf{x} \in \mathbb{R}^{dn}$ for n nodes. From this point on we'll assume we're working with linear tetrahedral elements in 3D, so we will have a single integration point per element and the integration weight just becomes the volume of the element, V_K. **(V_K is just a scalar - right? If so can we make it lower case?)** Using an implicit euler time integrator, our discrete variational form for the timestep $t + 1$ is

$$\mathbf{x}^{t+1}, \mathbf{s}^{t+1}, \lambda^{t+1} = \underset{\mathbf{x}, \mathbf{s}, \lambda}{\arg\min} \frac{1}{2h^2}||\mathbf{x} - \tilde{\mathbf{x}}||_M^2 + \sum_{K \in \mathcal{T}} \left(\Psi(\mathbf{s}_K) - \lambda_K \cdot (W_K\mathbf{s}_K - J_K\mathbf{x}) \right) V_K. \tag{2}$$

There's several changes here compared to the the continuous form so we explain them one-by-one. The first term, the kinetic energy term, is written now as a difference in the M-norm where M is the mass matrix for our mesh. Next we see that S is now a vector $\mathbf{s} \in \mathbb{R}^{6n_e}$ where $\mathbf{s}_K \in \mathbb{R}^6$ because there are only 6 degrees of freedom for a 3×3 symmetric matrix. Similarly our Lagrange multipliers are

now $\lambda \in \mathbb{R}^{9n_e}$ where each element's Lagrange multipliers are flattened into a vector λ_K. Another new piece is the matrix W_K which equals

$$W_K s_K = \text{vec}(R_K S_K), \tag{3}$$

and so it is constructed from R_K. W_K is of size 9×6 because it has the effect of performing a matrix multiplication between the rotation matrix and a symmetric matrix, then flattening the result. To simplify notation we just write W_K and R_K but to be more precise we would write $W_K(F_K(\mathbf{x}))$ and $R_K(F_K(\mathbf{x}))$ to indicate that the rotation is a function of the deformation gradient for the current positional configuration. **(More messy looking but I'd prefer to write those dependencies out explicitly in our energies to avoid bugs and misunderstanding downstream.)** Lastly we have rewritten the deformation gradient as

$$J_K \mathbf{x} = \text{vec}(F_K(\mathbf{x})) \tag{4}$$

where $J_K \in \mathbb{R}^{9 \times dn}$ is a Jacobian matrix.

Next to make our energy more suitable for an iterative Newton-like method, we note that

$$\mathbf{x}^{t+1} = \mathbf{x}^t + \delta\mathbf{x}$$
$$\mathbf{s}^{t+1} = \mathbf{s}^t + \delta\mathbf{s}$$
$$\lambda^{t+1} = \lambda^t + \delta\lambda,$$

so that we can instead solve for updates of our variables. Our discrete mixed energy becomes

$$
\begin{aligned}
E(\mathbf{x}^{t+1}, \mathbf{s}^{t+1}, \lambda^{t+1}) = {} & \frac{1}{2h^2} \|\mathbf{x}^t + \delta\mathbf{x} - \tilde{\mathbf{x}}\|_M^2 \\
& + \sum_{K \in \mathcal{T}} \Psi(\mathbf{s}_K^t + \delta\mathbf{s}_K) V_K \\
& - \sum_{K \in \mathcal{T}} (\lambda_K^t + \delta\lambda_K) \cdot (W_K(\mathbf{s}_K^t + \delta\mathbf{s}_K) - J_K(\mathbf{x}^t + \delta\mathbf{x})) V_K.
\end{aligned} \tag{5}
$$

and W_K in the final term is a function of $\mathbf{x}^t + \delta\mathbf{x}$.

To simplify things for derivations we split up the three terms in this energy to get

$$E(\mathbf{x}, \mathbf{s}, \lambda) = E_M(\mathbf{x}) + E_\Psi(\mathbf{s}) - E_\Lambda(\mathbf{x}, \mathbf{s}, \lambda), \tag{6}$$

and to arrive at a Newton step we just need to linearize with respect to the δ variables and construct our system of equations. The only hurdle to doing this is that E_Ψ is potentially high order, and E_Λ is nonlinear in $\mathbf{x}$ from the rotations, so we need to form quadratic approximations of these energies. **Last two sentences are a bit contradictory I think? Newton process is no more/no less than just forming a quadratic approximation (not linear) via taylor expansion and then solving it. Probably good for us to confirm we're thinking of the same thing here before we get into the nitty gritty :). Also, for creating a Newton solver there's no need to create this bunhc of extra δ terms in the above and below. Probably makes it messier than you need it to be for sanity checking.**

2 Energy Derivatives

In the previous section we split our discrete mixed energy intro three different terms, and before we write our final system of equations for each Newton step, we need to sort out the necessary derivatives and energy approximations.

2.1 Kinetic Energy

Or kinetic energy term for $\mathbf{x} = \mathbf{x}^t + \delta\mathbf{x}$ is

$$E_M(\mathbf{x}) = \frac{1}{2h^2}||x^t + \delta\mathbf{x} - \tilde{x}||_M^2 \tag{7}$$

$$= \frac{1}{2h^2}||\delta\mathbf{x} - h\dot{x}^t - h^2\mathbf{f}_{\text{ext}}||_M^2, \tag{8}$$

and its derivative with respect to $\delta\mathbf{x}$ is

$$\frac{\partial E_M}{\partial \delta\mathbf{x}} = \frac{1}{h^2}M(\delta\mathbf{x} - h\dot{\mathbf{x}}^t - h^2\mathbf{f}_{\text{ext}})$$

$$= \frac{1}{h^2}M(\delta\mathbf{x} - \mathbf{g}_\mathbf{x}^t) \tag{9}$$

2.2 Deformation Energy

The deformation energy is typically quadratic or some higher order function of $\mathbf{s}$, so we take a quadratic approximation of the energy, so that for $\mathbf{s} = \mathbf{s}^t + \delta\mathbf{s}$ we have

$$E_\Psi(\mathbf{s}) \approx \tilde{E}_\Psi(\mathbf{s}) = \sum_{K \in \mathcal{T}} \left(\Psi(\mathbf{s}_K^t) + \delta\mathbf{s}_K^T g_K + \frac{1}{2}\delta\mathbf{s}_K^T H_K \delta\mathbf{s}_K \right) V_K \tag{10}$$

where $\mathbf{g}_{\mathbf{s}K}^{t\,t} = \left.\frac{\partial\Psi}{\partial s_K}\right|_{s_K^t}$ and $H_K^t = \left.\frac{\partial^2\Psi}{\partial s_K^2}\right|_{s_K^t}$, and its derivative is simply

$$\frac{\partial\tilde{E}_\Psi}{\partial\delta\mathbf{s}_K} = \sum_{K \in \mathcal{T}} \left(g_K^t + H_K^t\delta\mathbf{s}_K \right) V_K \tag{11}$$

2.3 Constraint Energy

First we have the original constraint energy

$$E_\Lambda(\mathbf{x}, \mathbf{s}, \lambda) = \sum_{K \in \mathcal{T}} (\lambda_K \cdot (W_K\mathbf{s}_K - J_K\mathbf{x})) V_K, \tag{12}$$

where we have $\mathbf{x} = \mathbf{x}^t + \delta\mathbf{x}$, $\mathbf{s} = \mathbf{s}^t + \delta\mathbf{s}$, and $\lambda = \lambda^t + \delta\lambda$. And for its quadratic approximation, $\tilde{E}_\Lambda$ we have

$$\tilde{E}_\Lambda(\mathbf{x}, \mathbf{s}, \lambda) = E_\Lambda(\mathbf{x}^t, \mathbf{s}^t, \lambda^t) + \sum_{K \in \mathcal{T}} \left(\delta\mathbf{x}^T\frac{\partial E_\Lambda^K}{\partial\mathbf{x}} + \delta\mathbf{s}_K^T\frac{\partial E_\Lambda^K}{\partial\mathbf{s}_K} + \delta\lambda_K^T\frac{\partial E_\Lambda^K}{\partial\lambda_K} \right.$$

$$\left. + \delta\lambda_K^T\frac{\partial^2 E_\Lambda^K}{\partial\lambda_K\partial\mathbf{s}_K}\delta\mathbf{s}_K + \delta\lambda_K^T\frac{\partial^2 E_\Lambda^K}{\partial\lambda_K\partial\mathbf{x}}\delta\mathbf{x} + \delta\mathbf{x}^T\frac{\partial^2 E_\Lambda^K}{\partial\mathbf{x}\partial\mathbf{s}_K}\delta\mathbf{s}_K + \frac{1}{2}\delta\mathbf{x}^T\frac{\partial^2 E_\Lambda^K}{\partial\mathbf{x}^2}\delta\mathbf{x} \right) V_K \tag{13}$$

where each of the derivatives is evaluated at $\mathbf{x}^t, \mathbf{s}^t, \lambda^t$. The main difficulty in these derivatives is dealing with the derivative of the rotation term. We do this by defining the rotation in terms of the SVD of the deformation gradient,

$$R(\mathbf{x}) = U_K(\mathbf{x})V_K(\mathbf{x})^T \quad \text{where } F_K(\mathbf{x}) = U_K\Sigma_K V_K^T, \tag{14}$$

and

$$\frac{\partial R_K}{\partial F_{ij}} = \frac{\partial U_K}{\partial F_{ij}} V_K^T + U_K \frac{\partial V_K^T}{\partial F_{ij}}. \tag{15}$$

We now temporarily switch to indicial notation, and make K a superscript. Our derivative is a tensor $\frac{\partial R_{ij}}{\partial F_{kl}} \in \mathbb{R}^{3\times3\times3\times3}$. Since our degrees of freedom are all written as vectors, we redefine this tensor by flattening the first two indices, and modifying the last two dimensions so that it has the effect of multiplying by a symmetric matrix and flattening the result. This gives us the tensor $\hat{W}_{ijk}^K \in \mathbb{R}^{9\times9\times6}$. In the same way that we have

$$W_{ij}^K s_j^K = \text{vec}(R_{kl}^K S_{lm}^K), \tag{16}$$

we also have

$$\hat{W}_{ijk}^K s_k^K = \frac{\partial \text{vec}(R_{lm}^K S_{mn}^K)}{\partial F_{(kl)}}. \tag{17}$$

Now we write out each of the derivatives:

$$\frac{\partial E_\Lambda^K}{\partial \mathbf{x}_i} = J_{ij}(\hat{W}_{jkl}^K \mathbf{s}_l^K - I_{jk})\lambda_k^K \tag{18}$$

$$\frac{\partial E_\Lambda^K}{\partial \mathbf{s}_i^K} = W_{ij}^K \lambda_j^K \tag{19}$$

$$\frac{\partial E_\Lambda^K}{\partial \lambda_i^K} = W_{ij}^K \mathbf{s}_j^K - J_{ij}^K \mathbf{x}_j \tag{20}$$

$$\frac{\partial^2 E_\Lambda^K}{\partial \lambda_i^K \partial \mathbf{s}_j^K} = W_{ij}^K \tag{21}$$

$$\frac{\partial^2 E_\Lambda^K}{\partial \mathbf{x}_i \partial \lambda_k^K} = J_{ij}(\hat{W}_{jkl}^K \mathbf{s}_l^K - I_{jk}) \tag{22}$$

$$\frac{\partial^2 E_\Lambda^K}{\partial \mathbf{x}_i \partial \mathbf{s}_l^K} = J_{ij}\hat{W}_{jkl}^K \lambda_k^K \tag{23}$$

$$\frac{\partial^2 E_\Lambda^K}{\partial \mathbf{x}_i^2} = \text{i'll get around to it} \tag{24}$$

$$\tag{25}$$

where I_{jk} is a 9×9 identity matrix.

3 Energy Proxy

Let $c_K(\mathbf{x}, \mathbf{s}) = W_K \mathbf{s}_K - J_K \mathbf{x}$ where W_K is a function of $\mathbf{x}$. This means that $E_\Lambda^K(\mathbf{x}, \mathbf{s}, \lambda) = \lambda_K \cdot c_K(\mathbf{x}, \mathbf{s})$. We then have a quadratic approximation of the entire energy that takes the form

$$E(\mathbf{x}, \mathbf{s}, \lambda) \approx \tilde{E}(\mathbf{x}, \mathbf{s}, \lambda) = E_M(\mathbf{x}^t) + E_\Psi(\mathbf{s}^t) - E_\Lambda(\mathbf{x}^t, \mathbf{s}^t, \lambda^t)$$

$$+ (\mathbf{x} - \mathbf{x}^t)^T \left(\frac{\partial E_M}{\partial \mathbf{x}} - \frac{\partial \mathbf{c}}{\partial \mathbf{x}} \lambda^t \right)$$

$$+ (\mathbf{s} - \mathbf{s}^t)^T \left(\frac{\partial E_\Psi}{\partial \mathbf{s}} - \frac{\partial \mathbf{c}}{\partial \mathbf{s}} \lambda^t \right)$$

$$- (\lambda - \lambda^t)^T \mathbf{c}(\mathbf{x}^t, \mathbf{s}^t) \qquad \left(\mathbf{c}(\mathbf{x}^t, \mathbf{s}^t) = \left. \frac{\partial E_\Lambda}{\partial \lambda} \right|_{\mathbf{x}^t, \mathbf{s}^t, \lambda^t} \right)$$

$$- (\mathbf{x} - \mathbf{x}^t)^T \left(\frac{\partial}{\partial \mathbf{x}} \left(\frac{\partial \mathbf{c}}{\partial \mathbf{s}} \lambda^t \right) \right) (\mathbf{s} - \mathbf{s}^t)$$

$$- (\mathbf{x} - \mathbf{x}^t)^T \left(\frac{\partial \mathbf{c}}{\partial \mathbf{x}} \right) (\lambda - \lambda^t) - (\mathbf{s} - \mathbf{s}^t)^T \left(\frac{\partial \mathbf{c}}{\partial \mathbf{s}} \right) (\lambda - \lambda^t)$$

$$+ \frac{1}{2}(\mathbf{s} - \mathbf{s}^t)^T \left(\frac{\partial^2 E_\Psi}{\partial \mathbf{s}^2} \right) (\mathbf{s} - \mathbf{s}^t)$$

$$+ \frac{1}{2}(\mathbf{x} - \mathbf{x}^t)^T \left(\frac{\partial^2 E_M}{\partial \mathbf{x}^2} - \frac{\partial}{\partial \mathbf{x}} \left(\frac{\partial \mathbf{c}}{\partial \mathbf{x}} \lambda^t \right) \right) (\mathbf{x} - \mathbf{x}^t)$$

Next we cancel out some terms and also take the second derivative of $\mathbf{c}$ to be zero and get

$$\tilde{E}(\mathbf{x}, \mathbf{s}, \lambda) = E_M(\mathbf{x}^t) + E_\Psi(\mathbf{s}^t) + (\mathbf{x} - \mathbf{x}^t)^T \left(\frac{\partial E_M}{\partial \mathbf{x}} \right)$$

$$+ (\mathbf{s} - \mathbf{s}^t)^T \left(\frac{\partial E_\Psi}{\partial \mathbf{s}} \right) - \lambda^T \mathbf{c}(\mathbf{x}^t, \mathbf{s}^t)$$

$$- (\mathbf{x} - \mathbf{x}^t)^T \left(\frac{\partial}{\partial \mathbf{x}} \left(\frac{\partial \mathbf{c}}{\partial \mathbf{s}} \lambda^t \right) \right) (\mathbf{s} - \mathbf{s}^t)$$

$$- (\mathbf{x} - \mathbf{x}^t)^T \left(\frac{\partial \mathbf{c}}{\partial \mathbf{x}} \right) (\lambda) - (\mathbf{s} - \mathbf{s}^t)^T \left(\frac{\partial \mathbf{c}}{\partial \mathbf{s}} \right) (\lambda) \tag{26}$$

$$+ \frac{1}{2}(\mathbf{s} - \mathbf{s}^t)^T \left(\frac{\partial^2 E_\Psi}{\partial \mathbf{s}^2} \right) (\mathbf{s} - \mathbf{s}^t)$$

$$+ \frac{1}{2}(\mathbf{x} - \mathbf{x}^t)^T \left(\frac{\partial^2 E_M}{\partial \mathbf{x}^2} \right) (\mathbf{x} - \mathbf{x}^t),$$

and now using that $x - x^t = \delta\mathbf{x}$ and $\mathbf{s} - \mathbf{s}^t = \delta\mathbf{s}$:

$$\tilde{E}(\mathbf{x}, \mathbf{s}, \lambda) = E_M(\mathbf{x}^t) + E_\Psi(\mathbf{s}^t) + \delta\mathbf{x}^T \left(\frac{\partial E_M}{\partial \mathbf{x}}\right) + \delta\mathbf{s}^T \left(\frac{\partial E_\Psi}{\partial \mathbf{s}}\right) - \lambda^T \mathbf{c}(\mathbf{x}^t, \mathbf{s}^t)$$

$$- \delta\mathbf{x}^T \left(\frac{\partial}{\partial \mathbf{x}}\left(\frac{\partial \mathbf{c}}{\partial \mathbf{s}}\lambda^t\right)\right)\delta\mathbf{s} - \left(\delta\mathbf{x}^T \frac{\partial \mathbf{c}}{\partial \mathbf{x}} + \delta\mathbf{s}^T \frac{\partial \mathbf{c}}{\partial \mathbf{s}}\right)\lambda$$

$$+ \frac{1}{2}\delta\mathbf{s}^T \left(\frac{\partial^2 E_\Psi}{\partial \mathbf{s}^2}\right)\delta\mathbf{s} + \frac{1}{2}\delta\mathbf{x}^T \left(\frac{\partial^2 E_M}{\partial \mathbf{x}^2}\right)\delta\mathbf{x}. \tag{27}$$

$$= E_M(\mathbf{x}^t) + E_\Psi(\mathbf{s}^t) + \delta\mathbf{x}^T \mathbf{g}_\mathbf{x}^t + \delta\mathbf{s}^T \mathbf{g}_\mathbf{s}^t - \lambda^T \mathbf{c}(\mathbf{x}^t, \mathbf{s}^t)$$

$$- \delta\mathbf{x}^T \left(\frac{\partial}{\partial \mathbf{x}}\left(\frac{\partial \mathbf{c}}{\partial \mathbf{s}}\lambda^t\right)\right)\delta\mathbf{s} - \left(\delta\mathbf{x}^T \frac{\partial \mathbf{c}}{\partial \mathbf{x}} + \delta\mathbf{s}^T \frac{\partial \mathbf{c}}{\partial \mathbf{s}}\right)\lambda$$

$$+ \frac{1}{2}\delta\mathbf{s}^T H^t \delta\mathbf{s} + \frac{1}{2h^2}\delta\mathbf{x}^T M \delta\mathbf{x}.$$

4   System of Equations

Now to setup our system of equations, we differentiate our energy with respect to the δ-variables and set equal to zero. We switch back to vector notation and make K a subscript to simplify the notation.

4.1   Equation for $\delta\mathbf{x}$

Last we differentiate $\tilde{E}$ with respect to $\delta\mathbf{x}$ (we omit $\frac{\partial^2 \mathbf{c}_K}{\partial \mathbf{x}^2}\delta\mathbf{x}$ for now):

$$\frac{\partial \tilde{E}}{\partial \delta\mathbf{x}} = \frac{1}{h^2}M\delta\mathbf{x} + \mathbf{g}_\mathbf{x}^t - \sum_{K\in\mathcal{T}} \left(J_K^T(\hat{W}_K^t \cdot \lambda_K^t)\delta\mathbf{s}_K + J_K^T(\hat{W}_K^t \cdot \mathbf{s}_K^t - I)\lambda_K\right) V_K \tag{28}$$

4.2   Equation for $\delta\mathbf{s}$

We start by differentiating $\tilde{E}$ with respect to $\delta\mathbf{s}$:

$$\frac{\partial \tilde{E}^K}{\partial \delta\mathbf{s}_K} = \mathbf{g}_{\mathbf{s}K}^t + H_K^t \delta\mathbf{s}_K - W_K^{t\,T}\lambda_K - (\hat{W}_K^t \cdot \lambda_K^t)^T J_K \delta\mathbf{x} \tag{29}$$

where $(\hat{W}_K^t \cdot \lambda_K^t)$ performs a reduction over the middle index of $\hat{W}_{ijk}^K$. Setting this derivative equal to zero and solving for $\delta\mathbf{s}_K$ gives

$$\delta\mathbf{s}_K = H_K^{t\,-1} \left(W_K^{t\,T}\lambda_K + (\hat{W}_K^t \cdot \lambda_K^t)^T J_K \delta\mathbf{x} - \mathbf{g}_\mathbf{s}^t\right). \tag{30}$$

4.3   Equation for $\delta\lambda$

In the same fashion we differentiate $\tilde{E}$ with respect to $\delta\lambda$:

$$\frac{\partial \tilde{E}^K}{\partial \lambda_K} = -\mathbf{c}_K(x^t, \mathbf{s}^t) - \left(\delta\mathbf{x}^T \frac{\partial \mathbf{c}_K}{\partial \mathbf{x}} + \delta\mathbf{s}^T \frac{\partial \mathbf{c}_K}{\partial \mathbf{s}}\right)$$

$$= -\mathbf{c}_K(x^t, \mathbf{s}^t) - (\hat{W}_K^t \cdot \mathbf{s}_K^t - I)^T J_K \delta\mathbf{x} - W_K^t \delta\mathbf{s} \tag{31}$$

where $(\hat{W}_K^t \cdot \mathbf{s}_K^t)$ performs a reduction over the last index of $\hat{W}_{ijk}^K$.

4.4 System of Equations

To build the system of equations we define some variables to keep things neat:

$$F = -\sum_{K \in \mathcal{T}} P_K^{\mathbf{s}\,T} \left((\hat{W}_K^t \cdot \lambda_K^t)^T J_K^w \right) \tag{32}$$

$$G = J^w - \sum_{K \in \mathcal{T}} P_K^{\lambda\,T} \left((\hat{W}_K^t \cdot \mathbf{s}_K^t)^T J_K^w \right) \tag{33}$$

$$D = -\sum_{K \in \mathcal{T}} P_K^{\lambda\,T} W_K^t P_K^{\mathbf{s}} \tag{34}$$

where J^w is the volume weighted Jacobian, and $P_K^\lambda \in \mathbb{R}^{9 \times 9n_e}$ and $P_K^{\mathbf{s}} \in \mathbb{R}^{6 \times 6n_e}$ are selection matrics for the K-th set of Lagrange multipliers and deformation variables, respectively. Additionally H is the block diagonal Hessian for s^t. We form the system by setting each of the derivatives with respect to the δ-variables equal to zero and get

$$\begin{pmatrix} \frac{1}{h^2}M & F^T & G^T \\ F & H^t & D^T \\ G & D & 0 \end{pmatrix} \begin{pmatrix} \delta\mathbf{x} \\ \delta\mathbf{s} \\ \lambda \end{pmatrix} = \begin{pmatrix} -\mathbf{g}_{\mathbf{x}}^t \\ -\mathbf{g}_{\mathbf{s}}^t \\ \mathbf{c}(\mathbf{x}^t, \mathbf{s}^t) \end{pmatrix}. \tag{35}$$

Next, to get this thing in a KKT-like form we substitute the solution for the $\delta\mathbf{s}$ into the other equations. So we have

$$\delta\mathbf{s} = H^{t^{-1}} \left(-\mathbf{g}_{\mathbf{s}}^t - D^T\lambda - F\delta\mathbf{x} \right), \tag{36}$$

and sorry cause it's about to get a little ugly, but for the first set of equations we have

$$\frac{1}{h^2}M\delta\mathbf{x} + F^T H^{-1} \left(-\mathbf{g}_{\mathbf{s}}^t - D^T\lambda - F\delta\mathbf{x} \right) + G^T\lambda = -\mathbf{g}_{\mathbf{x}}^t$$

and collect terms to get

$$\left(\frac{1}{h^2}M - F^T H^{-1} F \right) \delta\mathbf{x} + (G^T - F^T H^{-1} D^T)\lambda = F^T H^{-1}\mathbf{g}_{\mathbf{s}}^t - \mathbf{g}_{\mathbf{x}}^t. \tag{37}$$

Next we substitute the solution for $\delta\mathbf{s}$ into the equations for λ:

$$G\delta\mathbf{x} + DH^{-1} \left(-\mathbf{g}_{\mathbf{s}}^t - D^T\lambda - F\delta\mathbf{x} \right) = \mathbf{c}(\mathbf{x}^t, \mathbf{s}^t)$$

and collect terms so that we have

$$(G - DH^{-1}F)\delta\mathbf{x} - DH^{-1}D^T\lambda = \mathbf{c}(\mathbf{x}^t, \mathbf{s}^t) + DH^{-1}\mathbf{g}_{\mathbf{s}}^t. \tag{38}$$

Before moving the equations back into matrix form, we define the following variables:

$$\tilde{M} = (\frac{1}{h^2}M - F^T H^{-1} F) \tag{39}$$

$$\tilde{J} = (G - DH^{-1}F). \tag{40}$$

Now our final system of equations for a single Newton step is

$$\begin{pmatrix} \tilde{M} & \tilde{J}^T \\ \tilde{J} & -DH^{-1}D^T \end{pmatrix} \begin{pmatrix} \delta\mathbf{x} \\ \lambda \end{pmatrix} = \begin{pmatrix} F^T H^{-1}\mathbf{g}_{\mathbf{s}}^t - \mathbf{g}_{\mathbf{x}}^t \\ \mathbf{c}(\mathbf{x}^t, \mathbf{s}^t) + DH^{-1}\mathbf{g}_{\mathbf{s}}^t \end{pmatrix}. \tag{41}$$

5 The various numerical methods

todo