Hyperelasticity

Conor Rowan

Winter 2024

Contents

1 Introduction

Elasticity in the regime of finite strains is confusing. The intention of this document is to help think through some core concepts such as different stress measures, the governing equations of nonlinear elasticity in the strong and weak forms, the idea that certain stress and strain measures are "work conjugate," and techniques for numerical solutions of the governing equations. The numerical results are also meant to demonstrate the power of using neural networks as a global discretization of the displacement. Using a global basis, energy minimization, and symbolic calculations makes it almost trivial to implement numerical

solutions to these nonlinear problems, which contrasts greatly with the serious implementation efforts needed with traditional finite element methods.
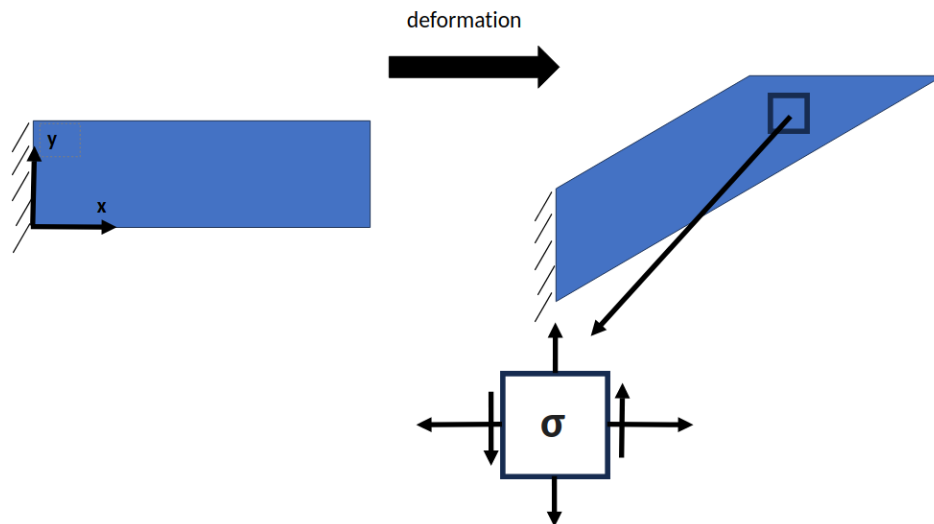
2   Different Stress Measures

Figure 1: The Cauchy stress tensor characterizes takes in a position in the deformed configuration and a normal vector and returns the traction vector on the plane at that point.

Force equilibrium is most natural to write in the deformed configuration. Nature does not care what geometry a structure had before the application of loads, rather only about the balance of forces at each point in a structure in whatever state it is currently in. Thus, we imagine deforming a structure, and characterizing its stress state by finding the traction vectors on three perpendicular faces aligned with the coordinate system at each point. This gives rise to the Cauchy stress σ_{ij}. See Figure 1. But because the goal of calculations in solid mechanics is often to find the deformation, quantities defined fully in the deformed configuration are somewhat vague and inconvenient. For example, we may want to know the traction/force on a plane defined by the geometry of the structure. The normal vector to this plane changes its orientation with the deformation, but its physical interpretation as a normal to some surface does not change. This motivates the idea of making use of the geometry of the undeformed configuration. The First Piola-Kirchhoff (PK1) stress tensor addresses exactly the question posed above: it takes in a point and normal vector in the undeformed configuration, and returns the traction on the plane at that point

from the deformed configuration. This allows us to query the stress state of the structure using the geometry of the reference configuration, which is more intuitive and convenient. For example in Figure 2, we may want the stress on the top surface of the bar in the undeformed configuration. If we use the PK1 stress, we do not need to account for the fact that this surface rotates 90 degrees under the action of forces when inputting a normal vector to compute this stress. We can use a result called "Nanson's formula" to relate the Cauchy and PK1 stresses. The deformation gradient $\underline{F}$ relates volumes in the two configurations with

$$dv = \det(F)dV$$

Capital letters are used to refer to the undeformed configuration, and lower case to the deformed configuration. The volume of a differential parallelpiped can be computed by dotting an area vector with another vector that defines its "height" (though it need not be normal to the area). Using the above relation, we can write

$$ds_i dx_i = JdS_i dX_i$$

where J is defined as the determinant of the deformation gradient. The area vectors can be written in the form $ds_i = dsn_i$ where n_i is the area normal and ds is the differential area. We can use this and the definition of the deformation gradient to write

$$dsn_i F_{ij} dX_j = JdSN_i dX_i$$

Factoring out the dX terms, this formula relates the area vectors in the deformed and undeformed configurations with

$$\underline{n}ds = J\underline{\underline{F}}^{-T}\underline{N}dS$$

By definition, the Cauchy stress takes in a normal and outputs the traction. This traction is turned into a differential force through a differential area. Thus, we have that

$$df_i = \sigma_{ij} n_j ds$$

Using Nanson's formula, we can write the differential area vector in the undeformed configuration. Thus we have the following equality:

$$df_i = \sigma_{ij} n_j ds = J\sigma_{ij} F_{kj}^{-1} N_k dS := P_{ik} N_k dS$$

This defines the PK1 stress in terms of the Cauchy stress and the deformation gradient. It takes in a normal vector in the undeformed configuration and spits out the traction on the plane defined by the deformed version of that normal. Note that the normal vector is in the undeformed configuration but the traction vector comes from the deformed configuration. Thus the direction of the force on the plane defined by the reference normal may seem odd and counter-intuitive:
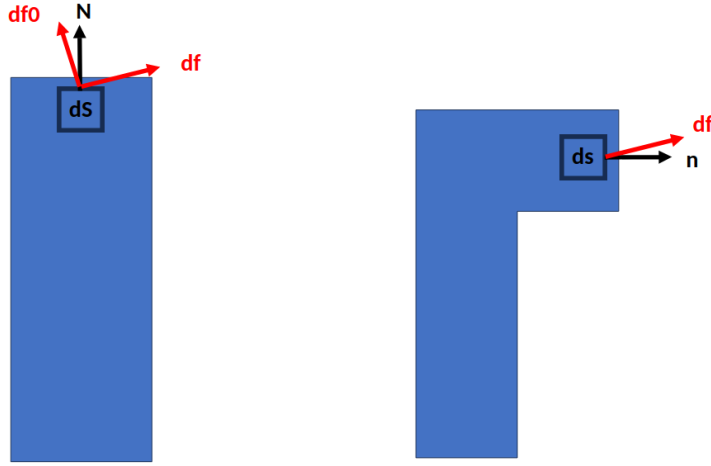
3

Figure 2: Reference configuration (left) and deformed configuration (right). A small bit of material with normal $\underline{N}$ and area dS transform under the deformation to the normal $\underline{n}$ and area ds. Under the action of the external loads, there is a force vector $d\underline{f}$ on the surface defined by $\underline{n}$. The fictitious force vector $d\underline{f}_0$ in the undeformed configuration has the same relation to the normal $\underline{N}$ as $d\underline{f}$ does to $\underline{n}$. It is clear that $\underline{n}$ and $\underline{N}$ refer to the same physical surface, but their components are quite different.

this is because the relation between the normal vector and the traction from the deformed state of the solid is not preserved. An element of material may be in pure tension in the deformed configuration but appear to have a component of stress parallel to the plane defined by the normal in the reference configuration. This is where the second Piola-Kirchhoff stress (PK2) can be useful–it is a stress measure which "pulls back" the force from the deformed configuration in the same way as the normal vectors are transformed. The traction vectors are mapped to the reference configuration with the deformation gradient in addition to the areas. The definition of the PK2 stress is

$$df_{0i} = JF_{ij}^{-1}\sigma_{jk}F_{\ell k}^{-1}N_\ell dS = S_{ij}N_j dS$$

where this differential force is a fictitious one rotated along with the normal vector back into the reference configuration. This is what the additional inverse of the deformation gradient is doing multiplying the PK1 stress.

3   Equilibrium Relations

The balance of linear momentum says that force is equal to the change in momentum of a body. For a region of material defined by the volume Ω, the

external forces come from tractions along its boundary and volumetric body forces. The statement of this principle for the body is

$$\frac{D}{Dt}\int_\Omega \rho v_i d\Omega = \int_{\partial\Omega} t_i dS + \int_\Omega \rho f_i d\Omega$$

The material time derivative is used to reflect the fact that we track a collection of particles. Remember that the balance of linear momentum applies not to a region in space, but to a particle or collection of particles. The Reynold's Transport theorem allows us to pass this time derivative inside the integral. Similarly, we can use the definition of the Cauchy stress, defined in the deformed configuration, to write this as

$$\int \rho \frac{Dv_i}{Dt} d\Omega = \int \sigma_{ij} n_j dS + \int \rho f_i d\Omega$$

We then use the divergence theorem to write the surface integral as a volume integral. Shrinking the volume Ω down to a point leads to the strong form of the equations of motion in nonlinear elasticity:

$$\int \left(\rho \frac{Dv_i}{Dt} - \frac{\partial \sigma_{ij}}{\partial x_j} - \rho f_i \right) d\Omega = 0$$

As above, formulation in the deformed configuration with the Cauchy stress is the most natural starting point. However, we want to determine the equations of mechanical equilibrium in the reference configuration in practice. We can convert each term in the balance of linear momentum to the reference configuration as follows:

$$\int_\Omega \rho f_i d\Omega = \int_{\Omega_0} \frac{\rho_0}{J} f_i J d\Omega_0 = \int_{\Omega_0} \rho_0 f_i d\Omega_0$$

$$\int_{\partial\Omega} t_i dS = \int_{\partial\Omega} \sigma_{ij} n_j dS = \int_{\partial\Omega_0} P_{ij} N_j dS_0 = \int_{\Omega_0} \frac{\partial P_{ij}}{\partial X_j} d\Omega_0$$

$$\frac{D}{Dt}\int_\Omega \rho v_i d\Omega = \frac{D}{Dt}\int_{\Omega_0} \rho_0 \frac{\partial u_i}{\partial t} d\Omega_0 = \int_{\Omega_0} \rho_0 \frac{\partial^2 u_i}{\partial t^2} d\Omega_0$$

The equations of motion in the reference configuration are then

$$\rho_0 \frac{\partial^2 u_i}{\partial t^2} = \frac{\partial P_{ij}}{\partial X_j} + \rho_0 f_i$$

Another way to see this in the case of statics is to start with the total potential energy in the reference configuration. With no body forces, this reads

$$\Pi = \int \Psi d\Omega_0 - \int t_i u_i dS_0$$

Note that another benefit of using the reference configuration is that it is more intuitive to specify the traction vectors. We want them to operate on

fixed surfaces of the undeformed body in given directions, not at spatial points. The displacement solution is obtained when the energy functional obtains a minimum. The calculus of variations tells us that this condition can be expressed as

$$\delta\Pi = \frac{\partial}{\partial\epsilon}\Pi(\underline{u} + \epsilon\underline{w})\Big|_{\epsilon=0} = 0$$

This says that small and arbitrary variations $\underline{w}$ around the displacement field do not change the value of the energy. This is the continuous generalization of the derivative of a function being zero. We think of the strain energy as being a function of the deformation gradient, which will return the PK1 stress. This is proven in the section of work conjugate pairs below. The point of using the PK1 stress is to define force equilibrium with the geometry of the reference configuration. The condition for a minimum is then

$$\delta\Pi = \int \frac{\partial\Psi}{\partial F_{ij}}(\underline{u}+\epsilon\underline{w})\frac{\partial}{\partial\epsilon}\left(\delta_{ij} + \frac{\partial u_i}{\partial X_j} + \epsilon\frac{\partial w_i}{\partial X_j}\right)\Big|_{\epsilon=0} d\Omega_0 - \int t_i\frac{\partial}{\partial\epsilon}(u_i+\epsilon w_i)\Big|_{\epsilon=0} dS_0 = 0$$

$$= \int P_{ij}\frac{\partial w_i}{\partial X_j}d\Omega_0 + \int t_i w_i dS_0$$

We have used that the PK1 stress and deformation gradient are work conjugate. Integrating this expression by parts, we arrive at

$$\int \frac{\partial P_{ij}}{\partial X_j}w_i d\Omega_0 - \int P_{ij}N_j w_i dS_0 + \int t_i w_i dS_0$$

Note that the variations are zero along the boundaries with prescribed displacements. If the displacement isn't prescribed, then the traction must be, which means that the two surface integrals in the above expression are over equivalent domains. Because the variation $\underline{w}$ is arbitrary, we can deduce the governing equation and its boundary conditions from this expression:

$$\frac{\partial P_{ij}}{\partial X_j} = 0 \quad \text{(no body force)}$$

$$P_{ij}N_j = t_i \quad \text{(on traction boundary)}$$

4   Work Conjugate Pairs

The work done by a displacement dependent force $F(x)$ over some deformation up to position x^* is

$$W = \int_0^{x^*} F(\xi)d\xi$$

For a time dependent displacement $x(t)$, the work done is a function of time through the displacement:

$$W(t) = \int_0^{x(t)} F(\xi)d\xi$$

By definition, power is the time derivative of the work. Because the time dependence of the work depends entirely on the limit of integration, we must use the Leibniz rule to compute this derivative. Note that the assumption of no explicit time dependence of the force rules out materials with time-dependent properties, such as the case of viscoelasticity. The power is computed as

$$p = \frac{\partial W}{\partial t} = \frac{\partial}{\partial t}\int_0^{x(t)} F(\xi)d\xi = \frac{\partial x}{\partial t}F(x(t))$$

This is where the familiar definition of force times velocity comes from. We can now extend this to elastic systems undergoing large strains. The total power of the body comes from external forces in the form of surface tractions $\underline{T}$ and body forces $\underline{f}$. The power now involves dot products of the force with the velocity:

$$p = \int T_i v_i dS + \int f_i v_i d\Omega$$

These integrals are taken in the current configuration $\Omega(t)$. The traction vector is related to the Cauchy stress with $T_i = \sigma_{ij}n_j$. Plugging this into the above expression and using the divergence theorem, the power can be written as

$$p = \int \frac{\partial}{\partial x_i}\left(\sigma_{ij}v_j\right)d\Omega + \int f_i v_i d\Omega$$

We can distribute the divergence, and use the governing equation of force equilibrium in the current configuration ($\sigma_{ij,j} + f_i = \rho a_i$) to arrive at

$$p = \frac{\partial}{\partial t}\int \frac{1}{2}\rho v_i v_i d\Omega + \int \sigma_{ij}\frac{\partial v_j}{\partial x_i}d\Omega$$

where we have used that the acceleration is the time derivative of the velocity. We can now recognize the first term and the time derivative of the total kinetic energy of the body, and the second term is the power associated with deformation. We will now focus on this term, naming it P (confusingly, because this is also the symbol for the PK1 stress). We can use the chain rule to rewrite the velocity gradient in a more revealing form:

$$P = \int \sigma_{ij}\frac{\partial v_i}{\partial x_j}d\Omega = \int \sigma_{ij}\frac{\partial v_i}{\partial X_k}\frac{\partial X_k}{\partial x_j}d\Omega = \int \sigma_{ij}\frac{\partial}{\partial t}\left(\delta_{ik} + \frac{\partial u_i}{\partial X_k}\right)\frac{\partial X_k}{\partial x_j}d\Omega$$

$$= \int \sigma_{ij}\dot{F}_{ik}F_{kj}^{-1}d\Omega$$

7

Now we perform a change of variables to integrate in the reference configuration. This means evaluating the Cauchy stress and deformation gradient at material points $\underline{X}$ as opposed to current positions $\underline{x}$. We also pick up a volume scaling factor:

$$P = \int \sigma_{ij} \dot{F}_{ik} F_{kj}^{-1} J d\Omega_0 = \int P_{ik} \dot{F}_{ik} d\Omega_0$$

We have shown that the power is computed with the PK1 stress and the time derivative of the deformation gradient. To finalize the demonstration that these two quantities together form the strain energy density (not just the power), we can use the fact that the total work done is the time integral of the power:

$$W = \int_0^T P(t)dt = \int_0^T \int P_{ik} \frac{\partial F_{ik}}{\partial t} d\Omega_0 dt$$

The PK1 stress does not have explicit time dependence, only depending on time implicitly through the deformation $\underline{\underline{F}}$. We know that for an elastic material, the total work done is independent of the path the deformation takes. Thus, we can compute the total work done in two ways: we can parameterize the deformation in terms of time, and sum the work contributions at each instant of time, or we can simply integrate the PK1 stress up to the final deformation $\underline{\underline{F}}(T)$. This amounts to a change of variables in the double integral, which reads

$$W = \int \left(\int_0^T P_{ik} \frac{\partial F_{ik}}{\partial t} dt \right) d\Omega_0 = \int \left(\int_0^{\underline{\underline{F}}(T)} P_{ik} dF_{ik} \right) d\Omega_0$$

This shows that the volumetric strain energy density Ψ is computed by integrating the PK1 stress against the deformation gradient. The relation that is typically more useful is when we start with an energy density (who knows where these come from!) and use it to compute the stress:

$$\frac{\partial \Psi}{\partial F_{ij}} = P_{ij}$$

That a stress measure is related to a strain measure through differentiating the energy is what is meant by work conjugacy. A similar argument using the power shows that the PK2 stress is work conjugate with the Green-Lagrange strain $\underline{\underline{E}}$.

5   Computational Hyperelasticity

Hyperelasticity is a large deformation constitutive model for which the deformation is reversible. This contrasts with more complex irreversible phenomena like damage and plasticity. The fact that the deformation is reversible means that problems in hyperelasticity can be formulated with a total potential energy functional. With no body force, this energy reads

$$\Pi = \int \Psi d\Omega_0 - \int t_i u_i dS_0$$

Integrals are carried out in the reference configuration. The response of the body to the applied traction is computed by finding the minimum of the energy functional. It is clear that the strain energy density Ψ governs the constitutive response of the body. We will use the compressible Neohookean model for 2D solids, which has a known strain energy density. This strain energy makes use of a few quantities. The right Cauchy-Green tensor and its first invariant are

$$C_{ij} = \underline{\underline{F}}^T \underline{\underline{F}} = F_{ki} F_{kj} \implies I_1 = \text{tr}(\underline{\underline{C}}) = C_{ii} = F_{ki} F_{ki}$$

The Neohookean model also makes use of the determinant of the deformation gradient

$$J = \det(\underline{\underline{F}}) = F_{11} F_{22} - F_{12} F_{21}$$

It is convenient that the determinant can be written in a compact and explicit form in two spatial dimensions. The strain energy density for a 2D incompressible Neohookean solid is

$$\Psi = C_1 \Big(I_1 - 2 - 2\ln J \Big) + D_1 \Big(J - 1 \Big)^2$$

where C_1 and D_1 are material constants. With the material model in hand, we are now in a position to say that the response of the body is governed by zero first variation (minimum) of the total potential energy:

$$\delta\Pi = \int \frac{\partial \Psi}{\partial F_{ij}} \delta F_{ij} d\Omega_0 - \int t_i \delta u_i dS_0 = 0$$

This is what is typically written in deriving the weak form of the governing equations. We differentiate the strain energy with respect to the deformation gradient because, as we have shown, this gives rise to the PK1 stress which allows us to work in the reference configuration. Because it is necessary to compute integrals over the body, it is essential that we operate in the reference configuration because otherwise the domain of integration would be unknown (as it depends on the deformation). The notation δF_{ij} is somewhat unclear–really what we mean by this is $F_{ij}(\delta \underline{u})$, i.e. the deformation gradient of the "variation" $\delta \underline{u}$. We will demonstrate the usual derivation of the weak form of the governing equations, but as will be seen, this is not the approach that will ultimately used in the solution process. This alternative approach is demonstrated in the following sections. Here, we discretize the variation as a linear combination of arbitrary coefficients multiplied by known spatial shape functions. Because the displacement is a vector, this needs to be done for each component separately. For example in two dimensions, we can write

$$\delta u_1 = \sum_{i=1}^{N} w_i N_i(\underline{X}), \quad \delta u_2 = \sum_{i=1}^{N} w_{i+N} N_i(\underline{X})$$

There are independent degrees of freedom for each component of the variation, but the spatial shape functions are the same. The discretization of the variation is done in the undeformed configuration, thus we use the coordinate $\underline{X}$ (upper case). We have $2N$ degrees of freedom total. It becomes very important to write the discretization in intuitive forms to keep the derivation manageable. The two above expressions can be combined into one in the following way:

$$
\begin{bmatrix} \delta u_1 \\ \delta u_2 \end{bmatrix} = \begin{bmatrix} N_1(\underline{X}) & N_2(\underline{X}) & \dots & N_N(\underline{X}) & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & N_1(\underline{X}) & \dots & N_N(\underline{X}) \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \\ w_{N+1} \\ \vdots \\ w_{2N} \end{bmatrix}
$$

This is a compact way to write the discretization of the variation in terms of a matrix of shape functions multiplied by a vector of all the degrees of freedom. Call the $2 \times 2N$ matrix of shape functions $H_{ij}(\underline{X})$ and the vector of degrees of freedom W_j. Then we have

$$
\delta u_i = H_{ij} W_j, \implies \delta F_{ij} = \frac{\partial \delta u_i}{\partial X_j} = \frac{\partial H_{ik}}{\partial X_j} W_k
$$

If you are wondering why $\delta F_{ij} \neq \delta_{ij} + \frac{\partial \delta u_i}{\partial X_j}$, which would be natural if we conceptualize the variation as a displacement around a position in the undeformed configuration, consider that in reality, what we are doing is the following:

$$
\delta \Psi = \frac{\partial \Psi}{\partial F_{ij}} \delta F_{ij} = \frac{\partial \Psi}{\partial F_{ij}} \frac{\partial F_{ij}}{\partial u_k} \delta u_k
$$

This expressions shows that the delta function does not show up. The weak form with the discretized variation is

$$
\Delta \Pi = 0 = \int \frac{\partial \Psi}{\partial F_{ij}} \frac{\partial H_{ik}}{\partial X_j} W_k d\Omega_0 - \int t_i H_{ij} W_j dS_0
$$

The coefficients on the variation are arbitrary, thus each term in the sum must be zero individually. Factoring out the coefficients and renaming indices, we obtain the residual system

$$
R_k = \int \frac{\partial \Psi}{\partial F_{ij}} \frac{\partial H_{ik}}{\partial X_j} d\Omega_0 - \int t_i H_{ik} dS_0 = 0
$$

At this point, we use a standard Galerkin method to approximate the displacement field in the same function space as the variation. This means that we use the same set of spatial shape functions. So as before, we have

$$
u_i = H_{ij}(\underline{X}) U_j
$$

10

We will also simplify the expression with $\partial \Psi / \partial F_{ij} = P_{ij}$. The first Piola-Kirchhoff stress is some nonlinear function of the displacement, which is defined by the strain energy density. Thus, the residual equations are a nonlinear system in terms of the displacement coefficients U_j. We write this as

$$R_k = \int P_{ij}(\underline{U}) \frac{\partial H_{ik}}{\partial X_j} d\Omega_0 - \int t_i H_{ik} dS_0 = 0$$

This is typically solved iteratively with a Newton-Raphson procedure. Each step in the Newton method involves linearizing the residual system around the current set of displacement coefficients $\underline{U}^n$, where n represents the current iteration. We find an update to the displacement coefficients such that linearized system is zero. In the vicinity of $\underline{U}^n$, the residual vector is approximated with a two term Taylor series

$$R_k(\underline{U}) \approx R_k(\underline{U}^n) + \frac{\partial R_k}{\partial U_j}\bigg|_{\underline{U}^n} \left(U_j - U_j^n\right) = 0$$

$$\implies \underline{U}^{n+1} = -\left(\frac{\partial \underline{R}}{\partial \underline{U}}\right)^{-1} \underline{R} + \underline{U}^n$$

The displacement at the next Newton iteration is the one for which the linearized residual system is solved. Because the system is not actually linear, this solution may or may not be accurate. For the hyperelastic problem, the Jacobian matrix needed for the Newton procedure is

$$\frac{\partial R_k}{\partial U_\ell} = \int \frac{\partial P_{ij}}{\partial U_\ell} \frac{\partial H_{ik}}{\partial X_j} d\Omega_0$$

Writing out the derivative of the PK1 stress with respect to the displacement coefficients would be a total mess. But the concept is quite simple, even if it is not simple in practice. These are the basic moving parts of a computational solution to a problem in hyperelasticity.

6 Ritz Method with Neural Network Basis for Hyperelasticity

A typical discretization is linear in the sense that the displacement is a linear combination of known spatial shape functions. The "parameters" to be determined are the weights on these shape functions. This is the technique that was employed above. However, a discretization is simply a tool for representing a wide class of functions. In two dimensions, the displacement is a function

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \to \begin{bmatrix} u_1(X_1, X_2) \\ u_2(X_1, X_2) \end{bmatrix}$$

The discretization is just some mathematical framework to capture this relationship. We think of the discretization as a means of expressing functions,

and the physics as a guide to tune this expression. Of course, any discretization is limited to a finite dimensional "subspace" of the space of all possible solutions. For example linear finite elements cannot represent curvature of the displacement field within elements. This is a restriction imposed by the choice of discretization. In recent years, neural networks have proven to be extremely flexible tools to represent complex input-output relationships. The simplest neural network architecture is a "multi-layer perceptron," which is a composition of linear transformations passed through nonlinear functions. For example, a single layer MLP neural network representing the displacement field can be written as

$$\underline{u} = \underline{\underline{w}}^2 \cdot \sigma\left(\underline{\underline{w}}^1\underline{x} + \underline{b}^1\right)$$

This is an affine transformation on the coordinates $\underline{x}$ defined by the "weight" matrix $\underline{\underline{w}}^1$ and "bias" vector $\underline{b}^1$, which is then passed through a nonlinear "activation" function $\sigma(\cdot)$. The activation is applied on a component-by-component basis. A matrix of coefficients then takes the output of the activation to the displacement. The weights and biases are all considered to be unknown parameters of this discretization. A deeper neural network repeats this process multiple times. For example, a two layer network looks like this:

$$\underline{u} = \underline{\underline{w}}^3\sigma\left(\underline{\underline{w}}^2 \cdot \sigma\left(\underline{\underline{w}}^1\underline{x} + \underline{b}^1\right) + \underline{b}^2\right)$$

This composition operation can be repeated arbitrarily many times. Because the weights at the last layer take spatial functions defined by the previous layers to the displacement, we might think of the network as approximating the shape functions in the inner layers, and the last layer as fitting their weights. In other words, if we freeze all the parameters other than those of the last layer, this looks like a typical linear discretization. The number of layers, the "widths" of the layers, and the activation functions are hyperparameters defining the structure of the network. A layers width is defined by the number of components being fed into the activation functions. Let's collect all the parameters into a single vector

$$\underline{\theta} = [\underline{\underline{w}}^1, \underline{b}^1, \underline{\underline{w}}^2, \dots]^T$$

The weight matrices can be reshaped into vectors for this to make sense. We will use a neural network with parameters θ to discretize the displacement field. This will be written as

$$\underline{u} = \underline{N}(X_1, X_2; \underline{\theta})$$

Using MATLAB, small to medium sized networks can be written out symbolically. We choose the number of layers, the width of each layer, the activation function, compute the total number of parameters, and build the weight and bias matrices by reshaping elements of the parameter vector. The functional relation between displacement and position is defined with the above composition operations symbolically. Unlike a finite element method, this discretization

is defined globally over the computational domain, as opposed to being local to elements. The benefit of symbolic calculations is tremendous. In the compressible Neohookean model, we require derivatives of the first invariant I_1 and the determinant of the deformation gradient J with respect to the parameters of the displacement discretization (for example in the Jacobian for the Newton solve). With the symbolic network, we can simply define these in terms of the displacement with

$$I_1 = I_1(\underline{C}) = I_1(\underline{C}(\underline{F})) = I_1(\underline{C}(\underline{F}(U))), \quad J = J(\underline{F}) = J(\underline{F}(U)))$$

and take derivatives in terms of the parameters symbolically. To make life even easier, we can use a Ritz method to solve for the hyperelastic deformation. Whereas the traditional weak form approach minimizes the energy with calculus of variations then discretizes the variation and displacement, the Ritz method starts with the energy, discretizes the displacement, then finds its minimum. If we use a gradient descent method to minimize the energy, there is no need to form a Jacobian matrix or use the Newton-Raphson method. Thus, all we need is the gradient of the discretized energy:

$$\frac{\partial \Pi}{\partial \theta_m} = \int \frac{\partial \Psi}{\partial \theta_m} d\Omega_0 - \int t_i \frac{\partial u_i}{\partial \theta_m} dS_0$$

For the compressible Neohookean model, the parameter gradient of the strain energy density is

$$\frac{\partial \Psi}{\partial \theta_m} = C_1 \left(\frac{\partial I_1}{\partial \theta_m} - \frac{2}{J} \frac{\partial J}{\partial \theta_m} \right) + 2D_1 \Big(J - 1 \Big) \frac{\partial J}{\partial \theta_m}$$

The gradient of the energy is used in an iterative search procedure to find a minimum in terms of the parameters θ. The parameter derivatives in the strain energy are computed symbolically as mentioned above. We write symbolic functions which evaluate the determinant of the deformation gradient, the first invariant of the right Cauchy-Green tensor and their derivatives at a given set of parameters. All we need to do at each step of the gradient descent algorithm is integrate these quantities over the computational domain.

7 Ritz Method with Neural Network Basis for Hyperelasticity with Contact

Some elastic contact problems are actually quite easy. This is especially true when using a Ritz method. We can simply add a penalty term to the energy functional that is large if the structure penetrates a surface. In general, the hard part is detecting penetration of surfaces for a given deformation. But for some problems, it is clear a priori where contact will occur, and simple to define penetration. In Figure 9, it is clear from the geometry and loading that the structure will contact the ground somewhere near the center. Penetration occurs
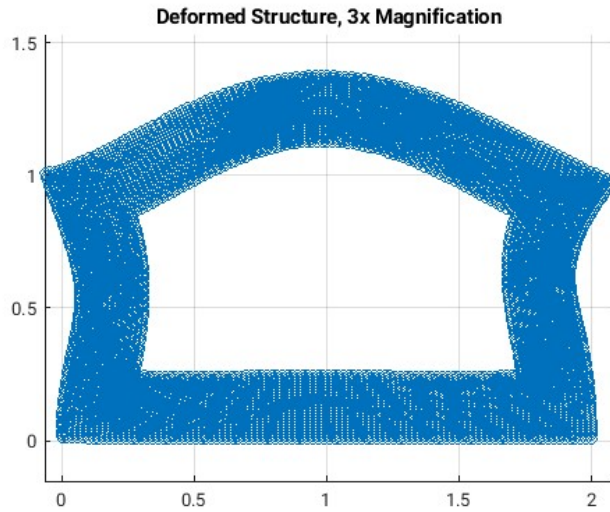
Figure 3: Frame structure with fixed lower surface and upwards traction applied at the center of the upper horizontal bar. A single hidden layer network with 10 neurons is used.
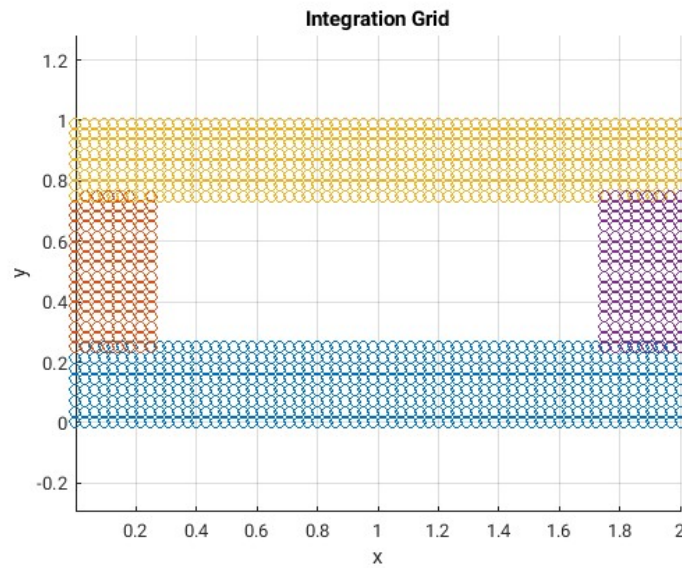

Figure 4: Integration grid for the frame structure.

when the position of any point on the lower surface of the bar passes through $y = 0$. This is very easy to express mathematically. We do this by defining
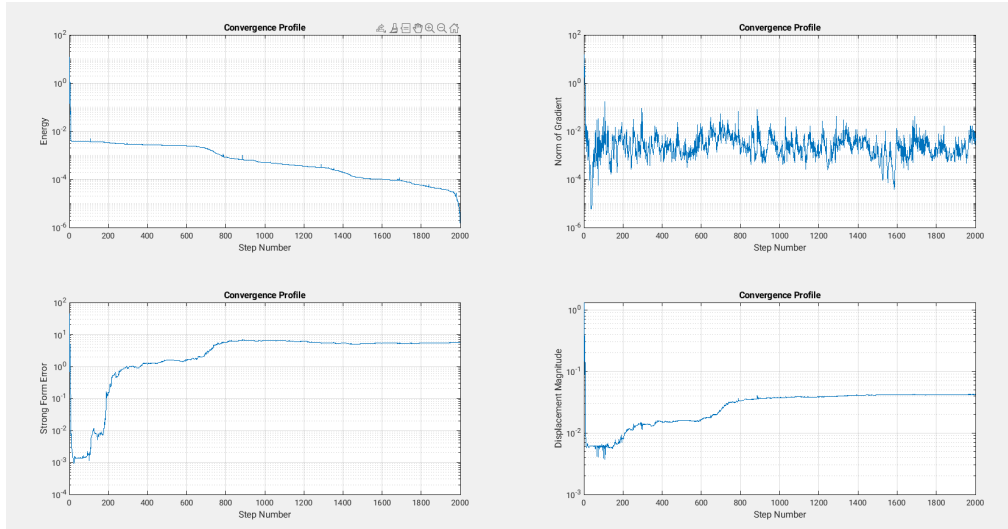
14

Figure 5: Convergence in the magnitude of the energy, the gradient, the strong form and the integral of the norm of the displacement field. The strong form error seems to track exactly with the size of the displacement. It seems that the optimizer initially makes the displacement small, then increases it which comes with a cost in the strong form error.

a "gap" function which measures the distance between the ground and points along the lower surface of the horizontal bar as a function of the displacement. For the arch structure in Figure 9, the gap can be written simply as

$$g(X_1) = L_2 - L_3 + u_2(X_1, L_2 - L_3)$$

where L_2 is the height of the arch from the ground, L_1 is the width of the arch, and L_3 is the width of the horizontal and vertical bars. Once we have the gap formulated, we can add a penalty to the energy functional of the form

$$\gamma \int_{L_3}^{L_1 - L_3} I^2\Big(g(X_1)\Big)dX_1$$

The parameter γ is a large penalty which has to be tuned. The function $I(g)$ is zero for $g > 0$, and $-g$ for $g < 0$. In other words, there is no penalty until penetration occurs. A function of this sort can be approximated as

$$I(g) = \frac{|g| - g}{2} \approx \frac{\sqrt{g^2 + \epsilon^2} - g}{2}$$

where ϵ is some (very) small parameter used to regularize the absolute value. The expression for the gap and the integration over the contact surface are specific to the arch problem, but the ideas are general: we need an expression
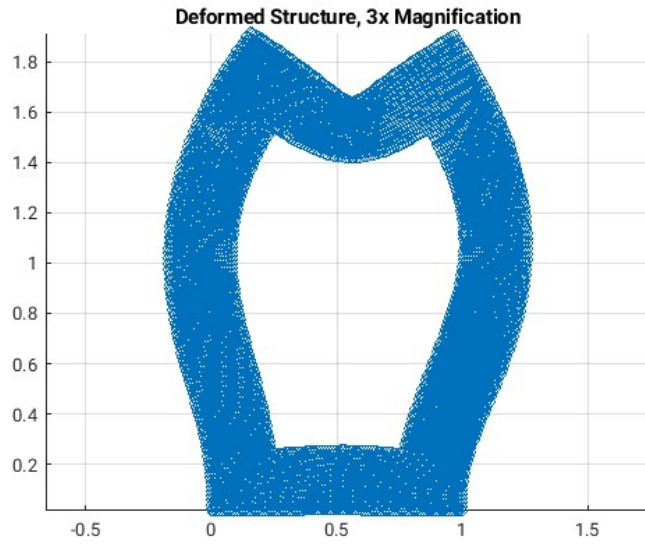
15

Figure 6: Frame structure with fixed lower surface and downwards traction applied at the center of the upper horizontal bar. A single hidden layer network with 10 neurons is used.
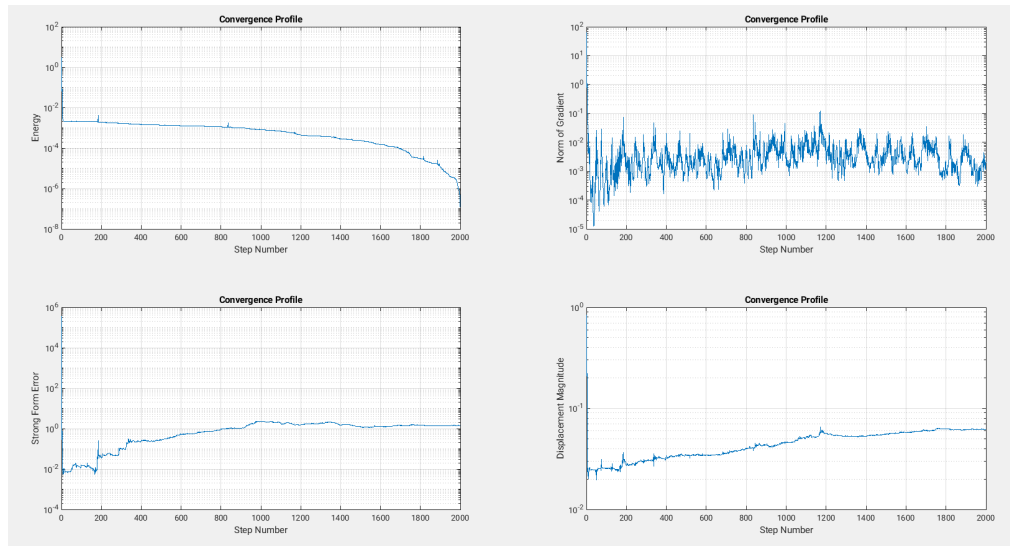

Figure 7: Convergence results for buckling frame structure.

for a gap function, and some area to integrate over for which contact may occur. This allows us to dramatically increase the value of the energy when penetration
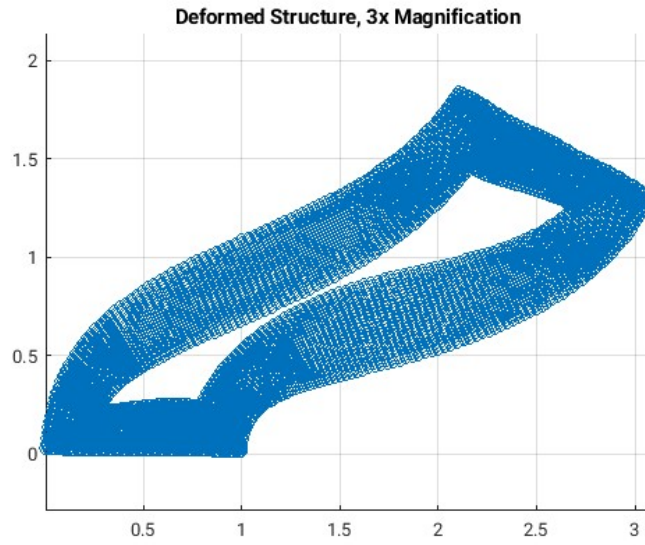
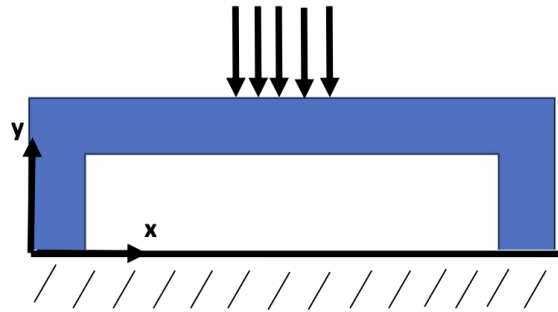Figure 8: Frame with applied shear traction on the upper surface.

Figure 9: Arch type structure which comes into contact with the ground under traction loading which is symmetric about the center.

occurs. The energy functional and its gradient for the hyperelastic arch with contact are

$$\Pi = \int \Psi d\Omega_0 - \int t_i u_i dS_0 + \gamma \int_{L_3}^{L_1-L_3} I^2\Big(g(X_1)\Big)dX_1$$

$$\frac{\partial \Pi}{\partial \theta_m} = \int C_1 \left(\frac{\partial I_1}{\partial \theta_m} - \frac{2}{J} \frac{\partial J}{\partial \theta_m} \right) + 2D_1 \left(J - 1 \right) \frac{\partial J}{\partial \theta_m} d\Omega_0 - \int t_i \frac{\partial u_i}{\partial \theta_m} dS_0$$
$$+ \gamma \int_{L_1}^{L_1 - L_3} 2I \left(g(X_1) \right) \frac{\partial I}{\partial g} \frac{\partial u_2}{\partial \theta_m} dX_1$$

See Figures 10-18 for results of numerical simulations for different geometries and loading.
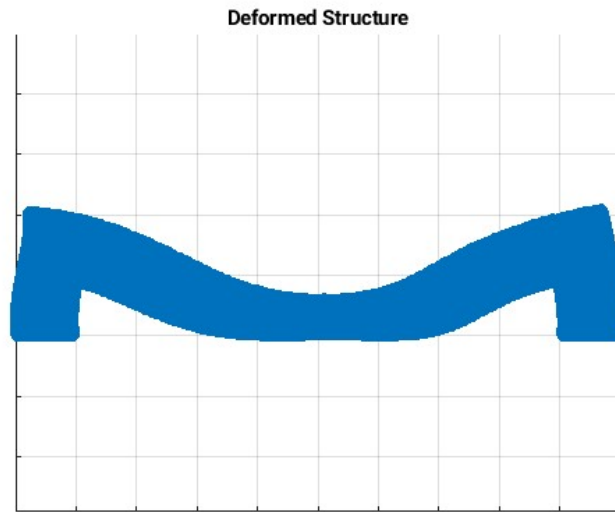

Deformed Structure

Figure 10: Arch with downward traction applied symmetrically about the center over one third the width. A single hidden layer network with 12 neurons was used.

8 Ritz Method with Neural Network Basis and Implicit Geometry

The geometry of a two-dimensional structure can be defined implicitly by the zero of some function g. The boundary is the curve

$$g(x_1, x_2) = 0$$

This is equivalent to thinking of the graph $x_3 = 0$ slice of the graph $x_3 = g(x_1, x_2)$. A function of this sort will be called a signed distance field, or level set function. A true signed distance field measures the shortest distance from
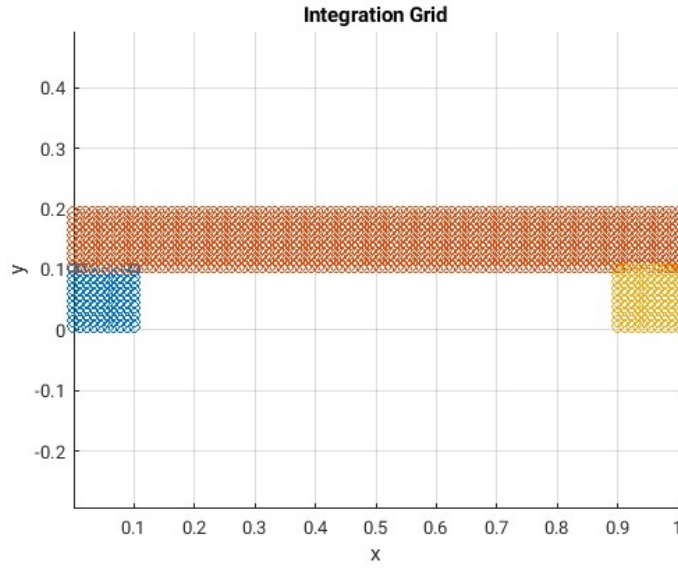
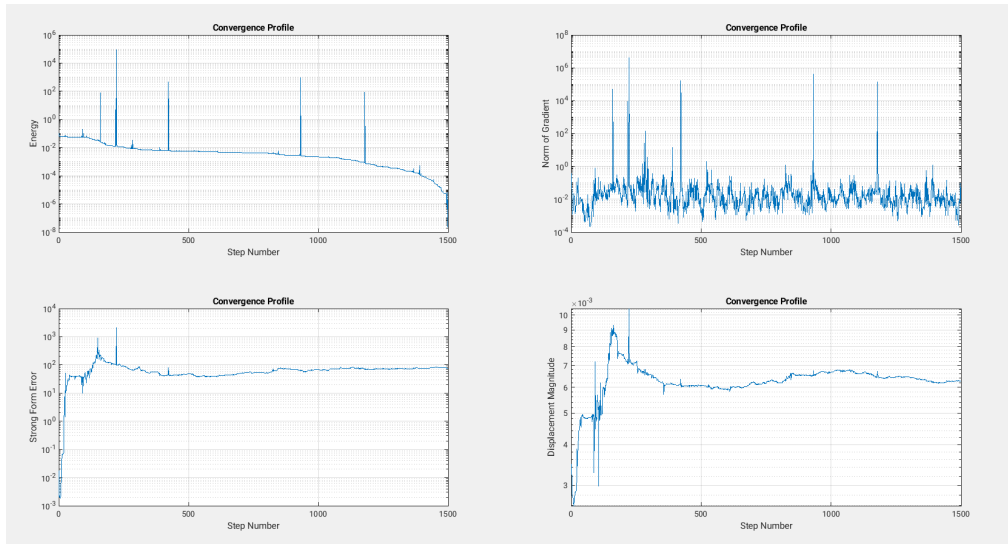Figure 11: Integration grid used in the arch contact problem.

Figure 12: Convergence profile of the energy, its gradient, the strong form error and the norm of the displacement field for the arch with centered traction loading.

a point in a domain to some curve defining a boundary. Some of the implicit geometries we construct may not be true SDF's, though the idea is the same.
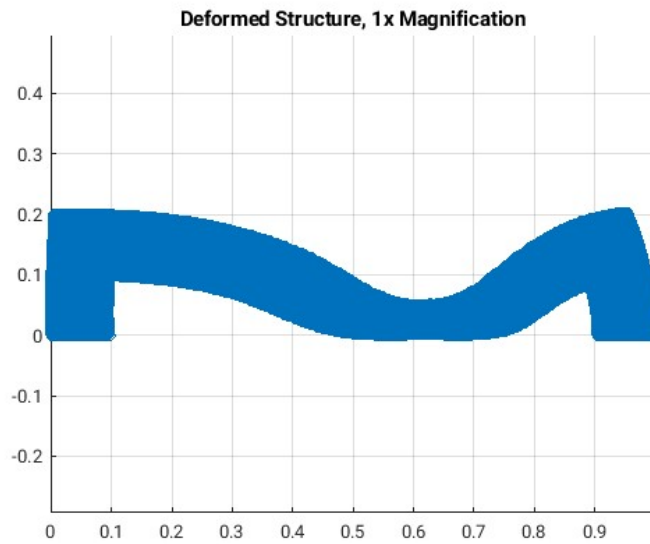
19

Figure 13: Arch with downward traction applied off center. The same neural network discretization and integration grid are used.
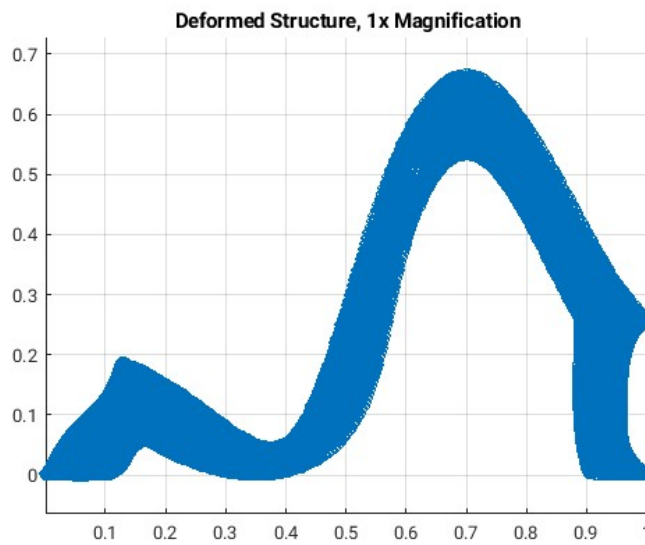

Figure 14: Arch with downward traction applied on the left and upward on the right.
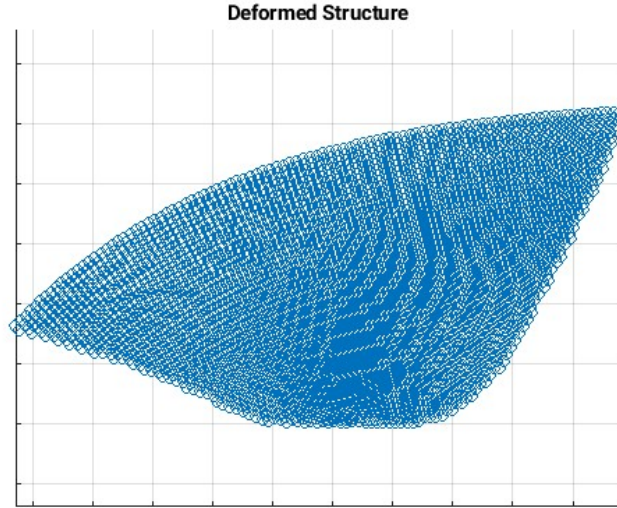
Figure 15: Parabola shaped structure with uniform vertical traction over the full width of the top surface. The structure topples to one side when symmetry of the displacement is not enforced explicitly.

It is some 3D function whose zero defines a boundary. If we have a simple SDF/level for circular, ellptical, or square geometry, we can build up more complex boundary by combining them with a max operation. See this plot for an illustration of this. For level sets $f_i(x_1, x_2)$, we can regularize the max function with

$$\max(f_i, \ldots, f_N)(x_1, x_2) \approx \frac{1}{\sum_{i=1}^{N} e^{\alpha f_i}} \sum_{i=1}^{N} f_i e^{\alpha f_i}$$

where α is some parameter which need not be particularly large ($\alpha > 10$ for example) for this regularization to be accurate. See Figures 19-22 for the solution of the hyperelasticity problem with zero boundaries and a body force using level set representation of the geometry using a neural network basis. We can simply multiply the neural network basis by the level set defined by the regularized max operation acting on four level sets defining ellipses in order to enforce the boundary conditions.
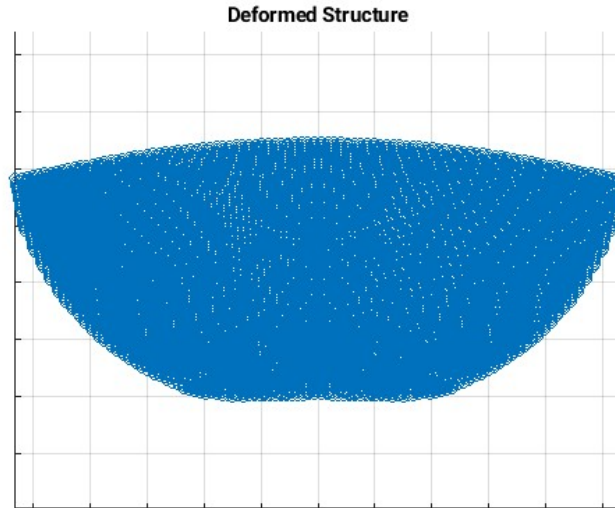
Deformed Structure

Figure 16: Deformation of hyperelastic hemisphere, computed using symmetry boundary conditions to avoid toppling. Zero x_1 displacement along the centerline is built into a single hidden layer network with 12 neurons. The size of integration elements is 1E-2 by 1E-2. The contact penalty parameter is 1E4 and the downward traction loading is applied uniformly over the upper surface. The solution is obtained after 700 optimization steps.

9   Ritz Method with Neural Network Basis for Multiscale Hyperelasticity

Analogous to the case of homogenization in linear elasticity, we can argue that for a two-scale hyperelastic material, the "macroscopic" stress is the average of the microscopic stress. The microscopic stress is the stress state of a block of material microstructure. The stress in the microstructure is driven by a macroscopic strain, which is a particular type of volumetric forcing. These are weird ideas until one has spent a lot of time thinking about multiscale problems. In the case of linear elasticity, these notions can be made rigorous with a two-scale perturbation expansion of the spatial coordinates and displacement field. See the Appendix for a discussion of this. The idea is that the material varies at high frequencies, which introduces the notion of separate scales. It can be seen with this plot that the displacement of a 1D bar with a constant body force and high frequency variations in the stiffness closely resembles a bar with the same body force but a reduced constant stiffness. In other words, the spatial variations primarily act to reduce the overall stiffness, not to introduce large oscillations in the solution. This suggests that "homogenized" material properties
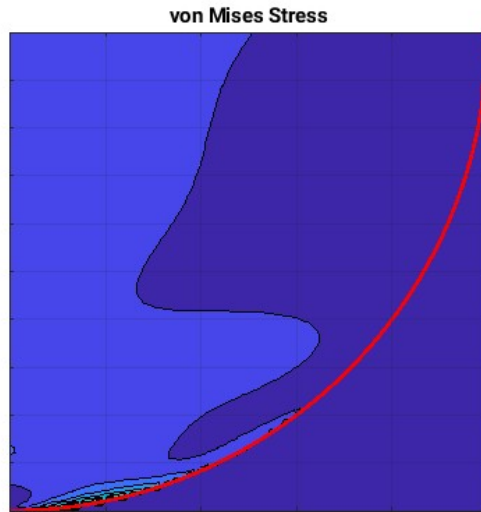
Figure 17: Von Mises stress for the symmetric part of the hemisphere, computed by converting the PK1 stress to the Cauchy stress. The largest stresses are along the contact surface.

Figure 18: Zooming in on large von Mises stresses along the contact surface.
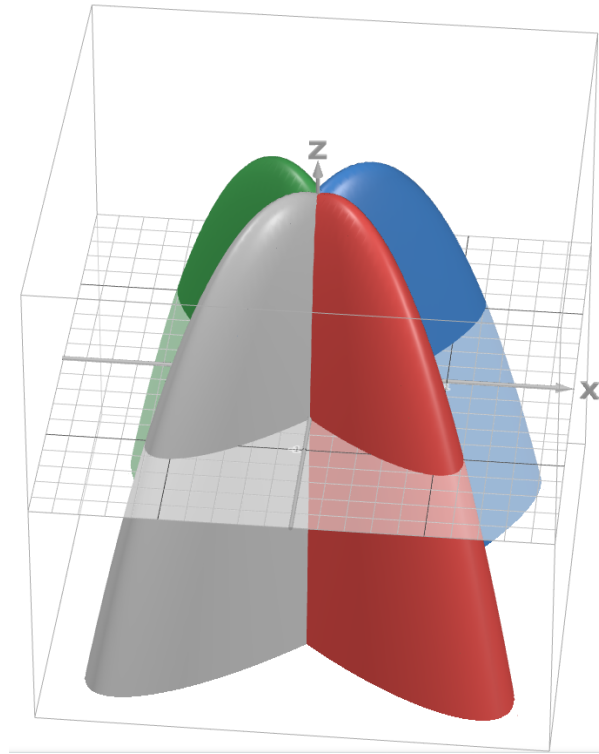
Figure 19: Signed distance field (SDF) aka level set representation of the geometry of a 2D structure. The zeros of the SDF implicitly define the boundary of the structure. Primitive SDF's can be combined through a "max" operation to construct more complex geometries.

can be a good model for a multiscale solid, and the two scale expansion is a way of estimating them. See the Appendix for a derivation of this, which shows that material can have macroscopic trends in addition to high frequency variations within the perturbation framework. In the linked plot, it can also be seen that directly averaging the material properties does not lead to an accurate model of the effective stiffness.

For linear problems, it seems fair to say that the perturbation techniques are the fundamental tool, and the interpretations of their results are basically heuristics to make some sense of things. Though it is possible for some nonlinear problems, it is not feasible in the case of hyperelasticity to honestly *derive* the multiscale model with the two scale perturbation expansion. One reason for this is that any nonlinearity ensures that the already tedious perturbation calculations become formidable. A more fundamental reason is that nonlinearities which are not polynomial make it impossible to "separate scales," which is
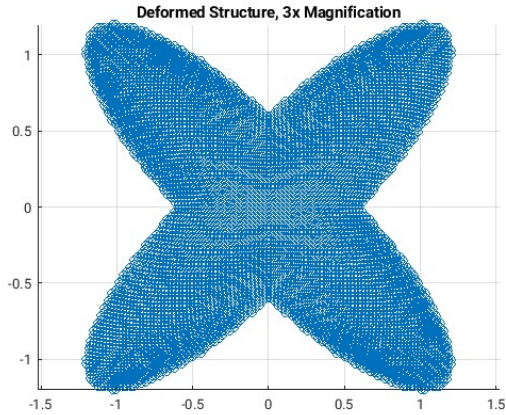
24

Figure 20: Geometry of 2D flower shaped hyperelastic structure constructed implicitly with a regularized max acting on four primitive level set functions.
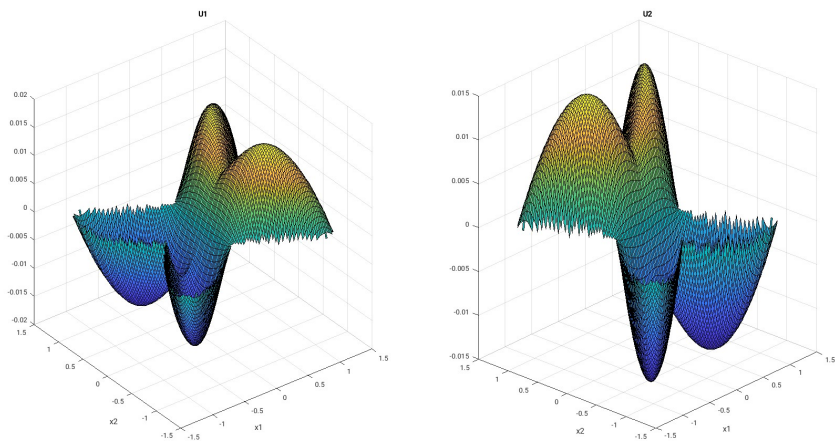
Figure 21: Displacement components for constant magnitude radial body force on structure whose geometry is defined with level sets. A single hidden layer neural network with 12 neurons is multiplied by the level set defining the geometry to enforce zero displacement boundary conditions. The size of integration elements is 3E-2 $\times$ 3E-2. Integration to form the energy and its gradient is now conducted with a background grid, where the level set function is used to check whether the integration point is inside or outside the boundary.

a necessary step in the usual homogenization procedure. Thus we rely on the intuition obtained from experience with other problems in order to propose a

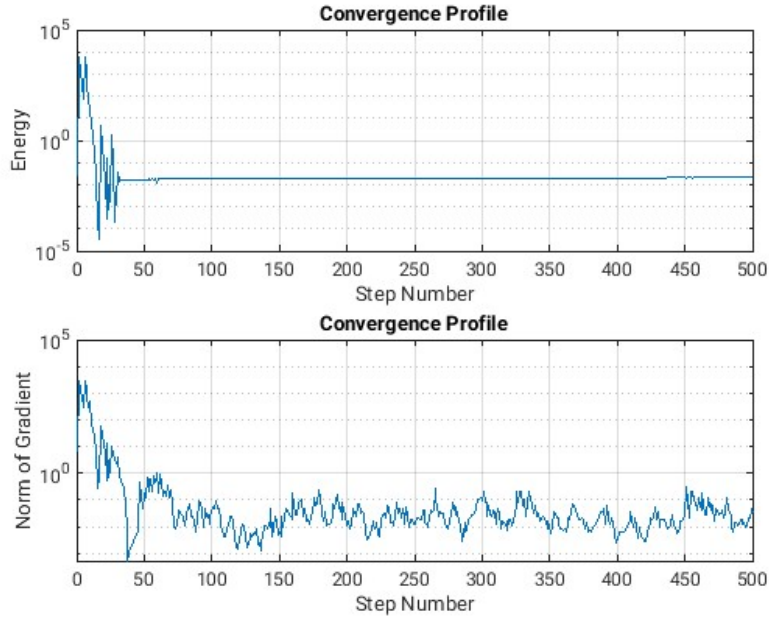Figure 22: Convergence of the energy and its gradient for flower shaped hyperelatsic structure. The radial body force seems to be a particularly simple solution to find.

hyperelastic multiscale model. So let's argue that the microstructure is hyperelastic with spatially varying material properties. A macroscopic deformation gradient, which is constant over the microstructure, acts as a forcing term for microscale equilibrium formulated in terms of the PK1 stress. We will use $\underline{Y}$ for the microscale coordinates in the reference configuration, and lower case letters for microscale quantities. Thus, $\underline{\underline{p}}$ is the microscale PK1 stress, and ψ is the microscale strain energy, which comes from knowledge of the material constituents of the microstructure. The governing equation reads

$$\frac{\partial p_{ij}}{\partial Y_j} = -\frac{\partial}{\partial Y_j}\frac{\partial \psi}{\partial f_{ij}}\bigg|_{\underline{\underline{F}}}$$

The microstructure has spatial variations, so the material parameters are no longer constant. The microscale hyperelastic material model for the same compressible Neohookean solid is

$$\psi = C_1(Y_1, Y_2)\Big(I_1 - 2 - 2\ln J\Big) + D_1(Y_1, Y_2)\Big(J - 1\Big)^2$$

When computing the stress response of the microstructure, the macroscopic deformation gradient is considered as a known quantity as it determines the

26

forcing term. It can be seen that this problem admits a variational formulation. The total potential energy corresponding to the microscale stress equilibrium is

$$\pi = \int \psi d\Omega^Y - \int u_i \frac{\partial}{\partial Y_j} \frac{\partial \psi}{\partial f_{ij}} \bigg|_{\underline{\underline{F}}} d\Omega^Y$$

The energy of the microstructure can be minimized for a given macroscopic deformation gradient in the same we solved the other hyperelasticity problems. The only difference is that the material is spatially varying, and that there is a peculiar body force. Solving for the displacement allows us to compute the microscopic PK1 stress, which then permits computation of the macroscale stress through an averaging operation:

$$P_{ij} = \frac{\partial}{\partial F_{ij}} \left(\frac{1}{|\Omega^Y|} \int \psi(Y_1, Y_2) d\Omega^Y \right) + \frac{1}{|\Omega^Y|} \int p_{ij}(\underline{\underline{F}}) d\Omega^Y$$
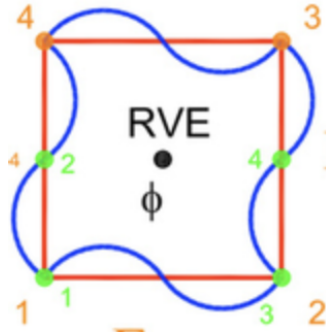

Figure 23: The simplest form of periodic boundary conditions where all four corners of the sample of microstructure are fixed.

This is the stress that comes from a deformation gradient acting on the direct average of the microstructure properties plus a correction from microscale stresses driven by material fluctuations. This is entirely analogous to linear homogenization (in fact the linear problem is the main motivation for this expression). The last thing that needs to be specified is the set of boundary conditions for the microstructure. We will use the simple form of periodic boundary conditions shown in Figure 23. Basically all multiscale models use periodic boundary conditions for the microstructure, but there are multiple ways they can be enforced. This is likely a "conservative" framework for enforcing periodic boundaries–we might think that shear type deformations should be possible. These are restricted by fixing the four corners. It is possible to build a neural network discretization of microscale displacement that which automatically satisfies this form of periodicity. This is written as

27

$$\underline{u}(Y_1, Y_2) = \sin(\pi Y_1)\sin(\pi Y_2)\underline{N}(Y_1, Y_2; \theta)$$
$$+ \left(e^{-Y_2^2} + e^{-(Y_2-1)^2}\right)\sin(\pi Y_1)\underline{g}(Y_1; \alpha) + \left(e^{-Y_1^2} + e^{-(Y_1-1)^2}\right)\sin(\pi Y_2)\underline{h}(Y_2; \beta)$$

See this plot for an illustration of a displacement field of this sort. The two vector valued functions of one spatial coordinate $\underline{h}$ and $\underline{g}$ represent the displacement along the boundaries. Here we are using that the size of the microstructure in terms of the microscale coordinate $\underline{Y}$ is unity. The parameters θ determine the solution inside the domain. The same displacement is imposed on opposite sides of the microstructure by construction. The half-period sine functions enforce zero displacement at the four corners of the microstructure. The inner displacement does not influence the boundaries because we multiply by the tensor product of sines. The boundary displacements decay to zero inside the domain. This discretization can be used with the microstructural energy minimization method to find the displacement, which is then used to compute the microscale stresses. From here, we can build a relationship between macroscopic deformation gradients and homogenized stresses.

A   Homogenization of 1D Bar

Say that there are periodic fluctuations in the stiffness of the bar with period η. This the parameter n in the plot. We conceptualize this parameter as specifying the size of the "microstructure" of the material. When η is very small, the entire microstructure appears to occupy a single point x. However, we can better resolve the microstructure by stretching out the position coordinate. This can be accomplished with the new coordinate $y := x/\eta$. Very small changes in x causes large changes in y. By definition, y changes value by 1 as a segment of microstructure is traversed. The two coordinates have the same order of magnitude. This is the set up for the perturbation techniques which are used to derive homogenized material properties. Looking at the analytical solution, we can see that when the size of the microstructure is small, even large amplitude fluctuations in the stiffness do not introduce large oscillations in the solution. To a good approximation, the bar responds as if there were a new value of constant material properties. We attempt to compute this "effective" stiffness. The first thing we do is treat x and y as independent coordinates. This is a strange idea given that they are obviously related, but the intuition is that when η is very small, the macroscopic position x is essentially constant as y varies within the microstructure. Think about this: if we define a spatially varying body force $b(x)$, when η is small, we have that $b(x + \eta) \approx b(x)$. This is all the effect that varying y has on the position x. It is thus reasonable, though very counterintuitive, to approximate these as two independent coordinates. Next, we argue that the effect of the periodic variations in the material is to introduce a small additive correction to an underlying displacement field. This displacement corrector is assumed to be periodic along with the material properties. The

superscript η is used to indicate a quantity which varies on two scales. The displacement is written as

$$u^\eta(x, y) = u_0(x) + \eta u_1(x, y)$$

The underlying displacement u_0 only varies with the macroscopic coordinate x. This is the "homogenized" displacement with updated but constant material properties we are trying to approximate. Note that if the stiffness has low and high frequency variations, the homogenization process will take out the high frequency variations, but there will still be some low frequency trend. This is shown in the linked plot. There is no reason to think that material properties varying with macroscopic coordinate x could be accurately replaced with a constant. We do think that the high frequency (microscale) fluctuations should be possible to ignore. We need to take derivatives of the displacement for stress equilibrium. Derivatives are written as

$$\frac{d}{dx}u^\eta = \frac{\partial u^\eta}{\partial x} + \frac{du^\eta}{dy} = \frac{\partial u^\eta}{\partial x} + \frac{\partial u^\eta}{\partial y}\frac{\partial y}{\partial x} = \frac{\partial u^\eta}{\partial x} + \frac{1}{\eta}\frac{\partial u^\eta}{\partial y}$$

The factor $1/\eta$ accounts for the fact that changes in the displacement with respect to the microscopic variable are actually happening much quicker than they seem when viewed from a zoomed in picture of the microstructure. This allows us to define the two-scale derivative operator as:

$$\frac{d}{dx^\eta} := \frac{\partial}{\partial x} + \frac{1}{\eta}\frac{\partial}{\partial y}$$

We assume that the stiffness of 1D bar varies with high and low frequency trends. The variations of the microstructure are periodic, but either the nature of this periodic variation or the mean around which the variation takes place change with x. Stress equilibrium for the two scale bar is written as

$$\frac{d}{dx^\eta}\left(E(x, y)\frac{du^\eta}{dx^\eta}\right) = b(x)$$

We have written the expression for the multiscale derivatives, and motivated the expansion of the two scale displacement as a homogenized displacement plus a small fluctuating correction. Using these results, stress equilibrium is expanded to read

$$\left(\frac{\partial}{\partial x} + \frac{1}{\eta}\frac{\partial}{\partial y}\right)\left(E(x, y)\left(\frac{\partial}{\partial x} + \frac{1}{\eta}\frac{\partial}{\partial y}\right)(u^0(x) + u^1(x, y))\right) = b(x)$$

Some algebra yields:

$$\frac{\partial}{\partial x}\left(E\frac{\partial u^0}{\partial x}\right) + \frac{1}{\eta}\frac{\partial}{\partial y}\left(E\frac{\partial u^0}{\partial x}\right) + \eta\frac{\partial}{\partial x}\left(E\frac{\partial u^1}{\partial x}\right) + \frac{\partial}{\partial y}\left(E\frac{\partial u^1}{\partial x}\right)$$
$$+ \frac{\partial}{\partial x}\left(E\frac{\partial u^1}{\partial y}\right) + \frac{1}{\eta}\frac{\partial}{\partial y}\left(E\frac{\partial u^1}{\partial y}\right) = b(x)$$

29

Now comes a peculiar trick from perturbation theory: we claim that equations need to be satisfied at each order of the small parameter η individually. I am not totally sure how this is justified, but one idea is that because η is so small, the size of the terms at different orders are completely mismatched. For example, terms involving $1/\eta$ will be extremely large, but it is not actually clear we want them to dominate the physics of the problem. In this case, it is really just that variations over the microstructure are extremely fast, so derivatives are large. This grouping by orders of η is a way of "separating scales." We look first at the terms on the lowest order of η, which in this case is η^{-1}. This equation reads

$$\frac{\partial}{\partial y}\left(E(x,y)\frac{\partial u^0}{\partial x}\right) = -\frac{\partial}{\partial y}\left(E(x,y)\frac{\partial u^1}{\partial y}\right)$$

We conceptualize this as the equation of the microscale. Imagine that the macroscopic displacement u^0 is known: what this equation says is that force equilibrium on the microscale (equation on u^1) is driven by a volumetric stress-like term arising from the macroscopic displacement field. This can be interpreted as the large scale driving a displacement (and stress) response on the level of the microstructure. This equation is linear, so we can write the displacement response of the microstructure at macroscopic position x as a function of the macroscopic strain with

$$u^1(x,y) = \chi(x,y)\frac{\partial u^0}{\partial x}$$

where χ is the microstructure response to a unit strain. This is nothing other than using linearity of this equation. Remember that x and y are treated as independent variables. The unit strain response of the microstructure depends on the macroscopic position x because the microstructure has a low frequency trend. As we will see, it will be very useful to write u^1 in terms of u^0. This is one of the things that the microscale equation has allowed us to do. Before moving forward, note that this equation can be solved analytically in 1D. The governing equation for the unit strain response is

$$\frac{\partial E(x,y)}{\partial y} = -\frac{\partial}{\partial y}\left(E(x,y)\frac{\partial \chi(x,y)}{\partial y}\right)$$

This can be integrated once to remove the derivatives, yielding

$$E(x,y)\frac{\partial \chi}{\partial y} = -E(x,y) + C$$

We can divide by E and integrate again to obtain

$$\chi(x,y) = -y + C\int_0^y \frac{d\xi}{E(x,\xi)} + D$$

The boundary conditions on the microscale unit strain response are periodic, given that it is used to define the displacement correction u^1. In 1D, this

30

means that $\chi(0) = \chi(1)$. As we will see, only derivatives of the unit strain response appear in the homogenized properties, which means that with no loss of generality we can take the displacement on the two boundaries to be zero. This requires that the constant of integration D is zero. The other constant of integration is determined with

$$\chi(x, 1) = 0 = -1 + C \int_0^1 \frac{d\xi}{E(x, \xi)}$$

$$\implies C = \frac{1}{\int_0^1 \frac{d\xi}{E(x,\xi)}}$$

The unit strain response of the microscale at macroscopic position x is then

$$\chi(x, y) = -y + C \int_0^y \frac{\partial\xi}{\partial E(x, \xi)}$$

where the constant C is defined by the above equation. Note that for finite scale separation, meaning $\eta > 0$, the microstructure defined by the interval $[E((x, 0), E(x, 1)]$ will "overlap" some amount with the microstructure on the interval $[E(x + \epsilon, 0), E(x + \epsilon, 1)]$ for any parameter $\epsilon < \eta$. It is not clear what the effect of this is. We can now move on to the macroscale equation. Turning to the second lowest order of η, we have

$$\frac{\partial}{\partial x}\left(E(x, y)\frac{\partial u^0}{\partial x}\right) + \frac{\partial}{\partial y}\left(E(x, y)\frac{\partial u^1}{\partial y}\right) + \frac{\partial}{\partial x}\left(E(x, y)\frac{\partial u^1}{\partial y}\right) = b(x)$$

Note that the body force is on the order η^0. What we can observe is the following: this equation cannot be satisfied pointwise in both x and y. The body force only depends on the macroscopic coordinate, but both the material and the displacement correction u^1 depend on the microstructure coordinate y. In other words, the body force on the right-hand side of the equation is constant over the microstructure, but the terms on the left-hand side vary. We can weaken the notion of a solution to be in an average sense only. In other words, the divergence terms on the left should equal the body force at point x when we average over the microstructure. It can be shown that this averaging operation comes from a technical condition on the equations at higher orders of η called "solvability." That being said, the intuitive interpretation seems reasonable. The average operation is

$$< \cdot > = \frac{1}{|\Omega^y|} \int (\cdot) d\Omega^y$$

where Ω^y is meant to indicate the volume of the microstructure. By definition, the measure of the microstructure domain (in this case length, not volume) is unity. This comes from the relation between the two coordinates. Applying the average operation, we have

$$\int \frac{\partial}{\partial x}\left(E(x,y)\frac{\partial u^0}{\partial x}\right) + \frac{\partial}{\partial y}\left(E(x,y)\frac{\partial u^1}{\partial y}\right) + \frac{\partial}{\partial x}\left(E(x,y)\frac{\partial u^1}{\partial y}\right)d\Omega^y = b(x)$$

It can be seen from periodicity of the material properties and the displacement corrector

$$\int \frac{\partial}{\partial y}\left(E(x,y)\frac{\partial u^1}{\partial y}\right)d\Omega^y = \left(E(x,y)\frac{\partial u^1}{\partial y}\right)\Big|_0^1 = 0$$

This is a convenient simplification of the macroscale governing equation. We can use the relation between u^1 and u^0 derived from the microscale equation to write this as

$$\int \frac{\partial}{\partial x}\left(E(x,y)\frac{\partial u^0}{\partial x}\right) + \frac{\partial}{\partial x}\left(E(x,y)\frac{\partial \chi}{\partial y}\frac{\partial u^0}{\partial x}\right)d\Omega^y = b(x)$$

The integration can be passed inside the macroscale derivatives. The equation we arrive at is

$$\frac{\partial}{\partial x}\left(\left[\int E(x,y)\left(1 + \frac{\partial \chi(x,y)}{\partial y}\right)d\Omega^y\right]\frac{\partial u^0}{\partial x}\right) = b(x)$$

This looks like stress equilibrium with updated material properties with no high frequency spatial variations. As noted earlier, there is still x dependence through both the stiffness $E(x,y)$ and the unit strain response $\chi(x,y)$. The influence of the microstructure in changing the effective material properties is accounted for with the unit strain response of the microstructure. We see that this offers a correction to directly averaging the stiffness over the microstructure. The new material properties can be interpreted as the direct average of the microstructure plus the average of a correction from microstructural stresses driven by the macroscopic strain. This is what the microstructural unit strain response is doing inside the homogenized tensor. In 1D, we can use the analytic expression for the unit strain response to show that the homogenized stiffness reduces to

$$\int E(x,y)\left(1 + \frac{\partial \chi(x,y)}{\partial y}\right)d\Omega^y = \int E(x,y)\left(1 - 1 + C\frac{1}{E(x,y)}\right)d\Omega^y$$

$$= C(x) = \frac{1}{\int_0^1 \frac{d\xi}{E(x,\xi)}}$$

This is a remarkably simple form. It is easy to see that the analytical solution for the homogenized bar is

$$u(x) = \int_0^x \frac{1}{C(y)}\int_0^y b(\xi)d\xi$$

32

When we don't include constants of integration, this displacement is zero with zero slope on the left end. This can be compared to the "true" analytical solution for the stiffness with high-frequency variations. Experimenting with animated plot shows that the homogenization framework works astonishingly well, even when the scale separation and periodicity assumptions are violated.

We saw that for the linearly elastic bar, the stress response was the stress arising from a direct average of the material properties on the microstructure, plus a correction term arising from the average of microscale stresses. We will argue that the homogenized response of the hyperelastic microstructure follows exactly this form: a direct average of the material properties plus a correction from the average of microstructural stresses. However, because hyperelasticity is nonlinear, we cannot "precompute" a unit strain response as a shortcut to the microstructural stresses, instead having to solve for the response of the microstructure for each applied deformation gradient.