U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

FFT-based Homogenization Methods

Professor:
Francisco Manuel Andrade Pires

Student:
José Luís Passos Vila-Chã

Report presented under the scope of the
Doctoral Program in Mechanical Engineering

Porto, September 2021

Page intentionally left blank.

Contents

List of Figures v

List of Tables vii

1 Continuum Mechanics and Finite Element Method 1
- 1.1 Kinematics of Deformation . . . 1
  - 1.1.1 Motion . . . 1
  - 1.1.2 Material and spatial descriptions . . . 2
  - 1.1.3 Deformation gradient . . . 2
    - 1.1.3.1 Isochoric/Volumetric decomposition . . . 3
    - 1.1.3.2 Polar decomposition . . . 3
- 1.2 Strain tensors . . . 4
- 1.3 Forces and stress measures . . . 5
  - 1.3.0.1 Cauchy stress tensor . . . 5
  - 1.3.0.2 First Piola-Kirchhoff stress tensor . . . 5
  - 1.3.0.3 Kirchhoff stress tensor . . . 6
  - 1.3.0.4 Deviatoric/Hydrostatic decomposition . . . 6
- 1.4 Heat . . . 6
  - 1.4.0.1 Heat flux vector . . . 6
- 1.5 Fundamental conservation principles . . . 6
  - 1.5.1 Principle of mass conservation . . . 7
  - 1.5.2 Principle of linear momentum conservation . . . 7
  - 1.5.3 First principle of thermodynamics . . . 7
  - 1.5.4 Second principle of thermodynamics . . . 8
  - 1.5.5 Clausius-Duhem inequality . . . 8
- 1.6 Thermomechanical constitutive initial value problem . . . 9
  - 1.6.1 Thermodynamics with internal variables . . . 10

2 Mechanical problem 13
- 2.0.1 Mechanical constitutive initial value problem . . . 13
- 2.0.2 Weak equilibrium. The principle of virtual work . . . 14
- 2.0.3 Mechanical constitutive initial boundary value problem . . . 15
- 2.1 Time descretization . . . 16
- 2.2 Finite Element Method . . . 18
  - 2.2.1 Finite element concept . . . 18
  - 2.2.2 Interpolation functions . . . 18
  - 2.2.3 Interpolation matrix and discrete gradient operators . . . 19
  - 2.2.4 Spatial discretization . . . 19

2.2.5 Numerical integration . . . 22

3 Thermal field **23**
3.1 Governing equations . . . 23
3.2 Thermal constitutive initial value problem . . . 23
3.3 Weak energy balance equation . . . 25
3.4 The thermal initial boundary value problem . . . 25
3.5 Finite Element Method . . . 27

4 Thermo-mechanical problem **31**
4.0.1 Thermo-mechanical constitutive initial value problem . . . 31
4.0.2 Weak equilibrium. The principle of virtual work . . . 33
4.0.3 Mechanical constitutive initial boundary value problem . . . 34
4.1 Time descretization . . . 36
4.2 Finite Element Method . . . 37
4.2.1 Interpolation . . . 38
4.2.2 Spatial discretization . . . 38

5 Validation results for the thermal solver **43**
5.1 Validation example 1 - DIN EN 1991-1-2/NA:2010-12: Anhang CC - Prüfung und Validierung von Rechenprogramm für Brandschutznachweise mittels allgemeiner Rechenverfahren - Beispiel 1) 43
5.1.1 Description . . . 43
5.1.2 Results . . . 45
5.2 Validation example 2 - DIN EN 1991-1-2/NA:2010-12: Anhang CC - Prüfung und Validierung von Rechenprogramm für Brandschutznachweise mittels allgemeiner Rechenverfahren - Beispiel 2) 48
5.2.1 Description . . . 48
5.2.2 Results . . . 48
5.3 Validation example 3 - The Standard NAFEMS Benchmarks: linear thermo-elastic tests - Two dimensional heat transfer with convection . . . 53
5.3.1 Description . . . 53
5.3.2 Results . . . 53

6 Solution procedures for coupled fields **57**
6.1 Context field elimination . . . 57
6.2 Monolithic . . . 58
6.2.1 Numerical considerations . . . 58
6.2.2 Usage examples . . . 60
6.3 Partitioned . . . 61
6.3.1 Operator splits . . . 62
6.3.2 Loosely vs. Strongly coupled schemes . . . 62
6.3.3 Loosely coupled . . . 63
6.3.4 Strongly coupled . . . 65
6.4 Comparison of solution techniques . . . 66

7 Implicit solution methods for coupled fields **71**
7.1 Equations to be solved . . . 71
7.2 A classification scheme for iterative methods . . . 72
7.2.1 Predictor . . . 73
7.2.2 Global Approaches . . . 74

7.2.3 Convergence criteria . . . 76
7.3 One-point iteration function . . . 77
7.3.1 Fixed-point approaches . . . 77
7.3.1.1 Block Jacobi or Schwarz additive . . . 77
7.3.1.2 Block Gauss-Seidel or Schwarz multiplicative . . . 77
7.3.2 Newton's method . . . 78
7.3.3 Constant Underrelaxation . . . 80
7.4 One-point iteration function with memory . . . 81
7.4.1 Aitken relaxation . . . 81
7.4.2 Multi-secant methods . . . 83
7.4.2.1 Generalized Broyden . . . 85
7.4.2.2 Anderson mixing . . . 86
7.4.2.3 Generalized Broyden's family . . . 86
7.4.2.4 Anderson's family . . . 87
7.4.2.5 The Broyden-like class . . . 87
7.4.3 Practical considerations . . . 88
7.5 Multipoint iteration functions . . . 93
7.5.1 Finite-Difference Newton Method . . . 93
7.5.2 Newton-Krylov methods . . . 94
7.5.3 Extrapolation tecniques in cycling mode . . . 96
7.6 Multipoint iteration functions with memory . . . 101
7.7 Summary . . . 101

8 Numerical results for the implicit coupling schemes 105
8.1 Expansion of a thermoelastic thick-walled cylinder . . . 105
8.1.1 Validation of the Numerical Results . . . 107
8.1.2 Evaluation and comparison of implicit solution methods for the coupled problem . . . 109
8.1.2.1 Methods with only one residual evaluation per iteration . 110
8.1.2.2 Broyden-like method . . . 112
8.1.2.3 Newton-GMRES method . . . 116
8.1.2.4 Polynomial vector extrapolation in cycling mode . . . 120
8.1.2.5 comparison of the best methods in each class . . . 122
8.1.2.6 Effect of predictiors . . . 126
8.2 Necking of a circular bar . . . 128
8.2.1 Validation of the Numerical Results . . . 128
8.2.2 Evaluation and comparison of implicit solution methods for the coupled problem . . . 133
8.2.2.1 Methods with only one residual evaluation per iteration . 133
8.2.2.2 Broyden-like method . . . 136
8.2.2.3 Newton-GMRES method . . . 138
8.2.2.4 Polynomial vector extrapolation in cycling mode . . . 141
8.2.2.5 comparison of the best methods in each class . . . 143
8.2.2.6 Effect of predictiors . . . 144
8.3 Conclusions . . . 146

List of Figures

1.1 Motion 2

2.1 Quasi-static mechanical constitutive initial boundary value problem. . . 15

5.1 Geometry and boundary conditions considered in the validation example 1 (**?**). 44
5.2 Numerical results for the validation example 1. (a) Temperature values at X as a function of time. (b) Relative error in percentage as function of time. 45
5.3 Numerical results regarding the evolution of the temperature distribution for the validation example 1 using a TRI3 mesh. 47
5.4 Geometry and boundary conditions considered in the validation example 2. (**?**) 49
5.5 Numerical results regarding the evolution of the temperature distribution for the validation example 2. (a) Temperature values at X as a function of time. (b) Relative error in percentage as function of time. 51
5.6 Numerical results for the validation example 2 using a TRI3 mesh. 52
5.7 Geometry and boundary conditions considered in the validation example 3 (**?**). 54
5.8 Temperature distribution concercing the validation example 3: (a) in two-dimensions (TRI6) (b) in three-dimensions (TETRA4). 56

6.1 Devices of partitioned analysis time-stepping (**?**). 63

7.1 Geometric interpretation of the fixed-point iteration method in one dimension. The fixed-point of f is sought, which is equivalent to the root of $x - f(x)$. 78
7.2 Geometric interpretation of the Newton method in one dimension for an example function f, whose derivative is denoted by f'. 79
7.3 Geometric interpretation of the Aitken relaxation in one dimension for an example function f and corresponding interpreation as the secant method. 82
7.4 Geometrical interpretation of Aitken's Δ^2 method. 97

8.1 Initial and boundary conditions considered in the quasi-static finite strain thermo-elastic expansion of an infinitely long thick-walled cylinder, and corresponding FEM mesh (QUAD4) used. 106

8.2 Difference between the temperature at the inner radius and the reference temperature for the expansion of the thick-walled cylinder with $\alpha_T = 1.65 \times 10^{-5}\,\mathrm{K}^{-1}$ and at different displacement rates for the inner radius ($\dot{u}_0 = 0.1\,\mathrm{mms}^{-1}$, $0.25\,\mathrm{mms}^{-1}$, and $0.5\,\mathrm{mms}^{-1}$ 108
8.3 Difference between the temperature at the inner radius and the reference temperature for the expansion of the thick-walled cylinder with $\alpha_T = 1.65 \times 10^{-4}\,\mathrm{K}^{-1}$ and at different displacement rates for the inner radius ($\dot{u}_0 = 0.1\,\mathrm{mms}^{-1}$, $0.25\,\mathrm{mms}^{-1}$, and $0.5\,\mathrm{mms}^{-1}$ 108
8.4 Temperature distribution in half a transversal section of the thick-walled cylinder at increasing inner radius displacements for $\alpha_T = 1.65 \times 10^{-4}\,\mathrm{K}^{-1}$ and $\dot{u}_0 = 0.5\,\mathrm{mms}^{-1}$. 109
8.5 Residual in percentage as a function of the number of nonlinear iterations in the first time step for the implicit methods that perform only one evaluation per nonlinear iteration in the solution of the quasi-static expansion of a thermoelastic thick-walled cylinder with $\alpha_T = 1.5 \times 10^{-4}\,\mathrm{K}^{-1}$ and $\dot{u}_0 = 0.5\,\mathrm{mms}^{-1}$. 110
8.6 number of nonlinear iterations needed to solve the coupled problem at each time step and the total number of iterations needed to solve the coupled problem for the implicit methods that perform only one evaluation per nonlinear iteration in the solution of the quasi-static expansion of a thermoelastic thick-walled cylinder with $\alpha_T = 1.5 \times 10^{-4}\,\mathrm{K}^{-1}$ and $\dot{u}_0 = 0.5\,\mathrm{mms}^{-1}$. 111
8.7 Total number of residual evaluations as a function of the thermal expansion coefficient for the implicit methods that perform only one evaluation per nonlinear iteration in the solution of the quasi-static expansion of a thermoelastic thick-walled cylinder with $\alpha_T = 0\,\mathrm{K}^{-1}$ to $1.5 \times 10^{-4}\,\mathrm{K}^{-1}$ and $\dot{u}_0 = 0.5\,\mathrm{mms}^{-1}$. 112
8.8 Residual in percentage as a function of the number of nonlinear iterations in the first time step for Broyden-like methods with Type I update and group sizes $s = 1$, 2, 4 and 6: (a) $\beta = -1$, (b) $\beta = 2 \times 10^{-3}$, and (c) $\beta = 2 \times 10^{-2}$ in the solution of the quasi-static expansion of a thermoelastic thick-walled cylinder with $\alpha_T = 1.5 \times 10^{-4}\,\mathrm{K}^{-1}$ and $\dot{u}_0 = 0.5\,\mathrm{mms}^{-1}$. 113
8.9 Residual in percentage as a function of the number of nonlinear iterations in the first time step for Broyden-like methods with Type II update and group sizes $s = 1$, 2, 4 and 6: (a) $\beta = -1$, (b) $\beta = 2 \times 10^{-3}$, and (c) $\beta = 2 \times 10^{-2}$ in the solution of the quasi-static expansion of a thermoelastic thick-walled cylinder with $\alpha_T = 1.5 \times 10^{-4}\,\mathrm{K}^{-1}$ and $\dot{u}_0 = 0.5\,\mathrm{mms}^{-1}$. 114
8.10 Number of nonlinear iterations needed to solve the coupled problem at each time step and the total number of iterations needed to solve the coupled problem for Broyden-like methods with Type I update and group sizes $s = 1$, 2, 4 and 6: (a) $\beta = -1$, (b) $\beta = 2 \times 10^{-3}$, and (c) $\beta = 2 \times 10^{-2}$ in the solution of the quasi-static expansion of a thermoelastic thick-walled cylinder with $\alpha_T = 1.5 \times 10^{-4}\,\mathrm{K}^{-1}$ and $\dot{u}_0 = 0.5\,\mathrm{mms}^{-1}$. 114
8.11 Number of nonlinear iterations needed to solve the coupled problem at each time step and the total number of iterations needed to solve the coupled problem for Broyden-like methods with Type II update and group sizes $s = 1$, 2, 4 and 6: (a) $\beta = -1$, (b) $\beta = 2 \times 10^{-3}$, and (c) $\beta = 2 \times 10^{-2}$ in the solution of the quasi-static expansion of a thermoelastic thick-walled cylinder with $\alpha_T = 1.5 \times 10^{-4}\,\mathrm{K}^{-1}$ and $\dot{u}_0 = 0.5\,\mathrm{mms}^{-1}$. 115

8.12 Total number of residual evaluations as a function of the thermal expansion coefficient for the implicit methods for Broyden-like methods with Type I update and group sizes $s = 1$, 2, 4 and 6: (a) $\beta = -1$, (b) $\beta = 2 \times 10^{-3}$, and (c) $\beta = 2 \times 10^{-2}$ in the solution of the quasi-static expansion of a thermoelastic thick-walled cylinder with $\alpha_T = 0\,\mathrm{K}^{-1}$ to $1.5 \times 10^{-4}\,\mathrm{K}^{-1}$ and $\dot{u}_0 = 0.5\,\mathrm{mms}^{-1}$. 115

8.13 Total number of residual evaluations as a function of the thermal expansion coefficient for Broyden-like methods with Type II update and group sizes $s = 1$, 2, 4 and 6: (a) $\beta = -1$, (b) $\beta = 2 \times 10^{-3}$, and (c) $\beta = 2 \times 10^{-2}$ in the solution of the quasi-static expansion of a thermoelastic thick-walled cylinder with $\alpha_T = 0\,\mathrm{K}^{-1}$ to $1.5 \times 10^{-4}\,\mathrm{K}^{-1}$ and $\dot{u}_0 = 0.5\,\mathrm{mms}^{-1}$. 116

8.14 Residual in percentage as a function of the number of nonlinear iterations in the first time step for the Newton-GMRES method with a constant forcing term ($\eta = 10^{-5}$, 10^{-3}, 10^{-1}) and the Eisenstat-Walker scheme in the solution of the quasi-static expansion of a thermoelastic thick-walled cylinder with $\alpha_T = 1.5 \times 10^{-4}\,\mathrm{K}^{-1}$ and $\dot{u}_0 = 0.5\,\mathrm{mms}^{-1}$. 117

8.15 Number of nonlinear iterations needed to solve the coupled problem at each time step and the total number of iterations needed to solve the coupled problem for the Newton-GMRES method with a constant forcing term ($\eta = 10^{-5}$, 10^{-3}, 10^{-1}) and the Eisenstat-Walker scheme in the solution of the quasi-static expansion of a thermoelastic thick-walled cylinder with $\alpha_T = 1.5 \times 10^{-4}\,\mathrm{K}^{-1}$ and $\dot{u}_0 = 0.5\,\mathrm{mms}^{-1}$. 118

8.16 Total number of residual evaluations as a function of the thermal expansion coefficient for the Newton-GMRES method with a constant forcing term ($\eta = 10^{-5}$, 10^{-3}, 10^{-1}) and the Eisenstat-Walker scheme in the solution of the quasi-static expansion of a thermoelastic thick-walled cylinder with $\alpha_T = 0\,\mathrm{K}^{-1}$ to $1.5 \times 10^{-4}\,\mathrm{K}^{-1}$ and $\dot{u}_0 = 0.5\,\mathrm{mms}^{-1}$. 119

8.17 Residual in percentage as a function of the number of nonlinear iterations in the first time step for the polynomial vector extrapolation methods in cycling mode, MPE and RRE, restricted to at most five evaluations of the residual function per nonlinear iteration in the solution of the quasi-static expansion of a thermoelastic thick-walled cylinder with $\alpha_T = 1.5 \times 10^{-4}\,\mathrm{K}^{-1}$ and $\dot{u}_0 = 0.5\,\mathrm{mms}^{-1}$. 120

8.18 number of nonlinear iterations needed to solve the coupled problem at each time step and the total number of iterations needed to solve the coupled problem for the polynomial vector extrapolation methods in cycling mode, MPE and RRE, restricted to at most five evaluations of the residual function per nonlinear iteration in the solution of the quasi-static expansion of a thermoelastic thick-walled cylinder with $\alpha_T = 1.5 \times 10^{-4}\,\mathrm{K}^{-1}$ and $\dot{u}_0 = 0.5\,\mathrm{mms}^{-1}$. 121

8.19 Total number of residual evaluations as a function of the thermal expansion coefficient for the polynomial vector extrapolation methods in cycling mode, MPE and RRE, restricted to at most five evaluations of the residual function per nonlinear iteration in the solution of the quasi-static expansion of a thermoelastic thick-walled cylinder with $\alpha_T = 0\,\mathrm{K}^{-1}$ to $1.5 \times 10^{-4}\,\mathrm{K}^{-1}$ and $\dot{u}_0 = 0.5\,\mathrm{mms}^{-1}$. 122

8.20 Total CPU time in seconds and the total number of residual evaluations as a function of the thermal expansion coefficient for the best performing implicit methods in each class considered in the solution of the quasi-static expansion of a thermoelastic thick-walled cylinder with $\alpha_T = 0\,\mathrm{K}^{-1}$ to $1.5\times10^{-4}\,\mathrm{K}^{-1}$ and $\dot{u}_0 = 0.5\,\mathrm{mms}^{-1}$. 124

8.21 Total CPU time in seconds, and time profile, as a function of the mesh size for the best performing implicit methods in each class considered in the solution of the quasi-static expansion of a thermoelastic thick-walled cylinder with $\alpha_T = 1.5\times10^{-4}\,\mathrm{K}^{-1}$ and $\dot{u}_0 = 0.5\,\mathrm{mms}^{-1}$. 125

8.22 Total number of iterations as a function of the thermal expansion coefficient for the best performing implicit methods in each class considered using a linear, a quadratic, and no predictor in the solution of the necking of a circular bar with $\alpha_T = 0\,\mathrm{K}^{-1}$ to $1.5\times10^{-4}\,\mathrm{K}^{-1}$ and $\dot{u}_0 = 0.5\,\mathrm{mms}^{-1}$. 127

8.23 Description of the thermally triggered necking of a circular bar problem, characteristic deformation, and temperature field stages during the loading and example axisymmetric finite element mesh. The results have been obtained with QUAD4 elements, non-adiabatic boundary conditions, the inconsistent mechanical dissipation formulation, and the Fourier law based on constant k_0. 130

8.24 evolution of the reaction force at the tips of the bar and the neck surface temperature with the prescribed displacement using QUAD8 elements with reduced integration (QUAD8R) and HEXA8-FBAR elements. 132

8.25 Total number of residual evaluations as a function of the thermal expansion coefficient for the fixed-point method in the solution of the necking of a circular bar with $w_0 = w_h = 2\times10^{-3}\,\mathrm{K}^{-1}$, $4.89\times10^{-3}\,\mathrm{K}^{-1}$, and $10^{-2}\,\mathrm{K}^{-1}$. 134

8.26 number of nonlinear iterations needed to solve the coupled problem at each time step and the total number of iterations needed to solve the coupled problem for the implicit methods that perform only one evaluation per nonlinear iteration in the solution of the necking of a circular bar with $w_0 = w_h = 10^{-2}\,\mathrm{K}^{-1}$. 135

8.27 Total number of residual evaluations as a function of the thermal expansion coefficient for the implicit methods that perform only one evaluation per nonlinear iteration in the solution of the necking of a circular bar with $w_0 = w_h = 2\times10^{-3}\,\mathrm{K}^{-1}$ to $10^{-2}\,\mathrm{K}^{-1}$. 135

8.28 Number of nonlinear iterations needed to solve the coupled problem at each time step and the total number of iterations needed to solve the coupled problem for Broyden-like methods with Type I update and group sizes $s = 1$, 2, 4 and 6: (a) $\beta = -1$, (b) $\beta = 2\times10^{-3}$, and (c) $\beta = 2\times10^{-2}$ in the solution of the necking of a circular bar with $w_0 = w_h = 10^{-2}\,\mathrm{K}^{-1}$. . . . 136

8.29 Number of nonlinear iterations needed to solve the coupled problem at each time step and the total number of iterations needed to solve the coupled problem for Broyden-like methods with Type II update and group sizes $s = 1$, 2, 4 and 6: (a) $\beta = -1$, (b) $\beta = 2\times10^{-3}$, and (c) $\beta = 2\times10^{-2}$ in the solution of the necking of a circular bar with $w_0 = w_h = 10^{-2}\,\mathrm{K}^{-1}$. . . . 137

8.30 Total number of residual evaluations as a function of the thermal expansion coefficient for the implicit methods for Broyden-like methods with Type I update and group sizes $s = 1$, 2, 4 and 6: (a) $\beta = -1$, (b) $\beta = 2 \times 10^{-3}$, and (c) $\beta = 2 \times 10^{-2}$ in the solution of the necking of a circular bar with $w_0 = w_h = 2 \times 10^{-3}\,\mathrm{K}^{-1}$ to $10^{-2}\,\mathrm{K}^{-1}$. 137
8.31 Total number of residual evaluations as a function of the thermal expansion coefficient for Broyden-like methods with Type II update and group sizes $s = 1$, 2, 4 and 6: (a) $\beta = -1$, (b) $\beta = 2 \times 10^{-3}$, and (c) $\beta = 2 \times 10^{-2}$ in the solution of the necking of a circular bar with $w_0 = w_h = 2 \times 10^{-3}\,\mathrm{K}^{-1}$ to $10^{-2}\,\mathrm{K}^{-1}$. 138
8.32 Number of nonlinear iterations needed to solve the coupled problem at each time step and the total number of iterations needed to solve the coupled problem for the Newton-GMRES method with a constant forcing term ($\eta = 10^{-5}$, 10^{-3}, 10^{-1}) and the Eisenstat-Walker scheme in the solution of the necking of a circular bar with $w_0 = w_h = 10^{-2}\,\mathrm{K}^{-1}$. . . . 139
8.33 Total number of residual evaluations as a function of the thermal expansion coefficient for the Newton-GMRES method with a constant forcing term ($\eta = 10^{-5}$, 10^{-3}, 10^{-1}) and the Eisenstat-Walker scheme in the solution of the necking of a circular bar with $w_0 = w_h = 2 \times 10^{-3}\,\mathrm{K}^{-1}$ to $10^{-2}\,\mathrm{K}^{-1}$. 140
8.34 number of nonlinear iterations needed to solve the coupled problem at each time step and the total number of iterations needed to solve the coupled problem for the polynomial vector extrapolation methods in cycling mode, MPE and RRE, restricted to at most five evaluations of the residual function per nonlinear iteration in the solution of the necking of a circular bar with $w_0 = w_h = 10^{-2}\,\mathrm{K}^{-1}$. 141
8.35 Total number of residual evaluations as a function of the thermal expansion coefficient for the polynomial vector extrapolation methods in cycling mode, MPE and RRE, restricted to at most five evaluations of the residual function per nonlinear iteration in the solution of the necking of a circular bar with $w_0 = w_h = 2 \times 10^{-3}\,\mathrm{K}^{-1}$ to $10^{-2}\,\mathrm{K}^{-1}$. 142
8.36 Total CPU time in seconds and the total number of residual evaluations as a function of the thermal softening parameters $w_0 = w_h$ for the best performing implicit methods in each class considered in the solution of the necking of a circular bar with $w_0 = w_h = 2 \times 10^{-3}\,\mathrm{K}^{-1}$ to $10^{-2}\,\mathrm{K}^{-1}$. . . . 143
8.37 Total number of iterations as a function of the thermal softening parameters $w_0 = w_h$ for the best performing implicit methods in each class considered using a linear, a quadratic, and no predictor in the solution of the necking of a circular bar with $w_0 = w_h = 2 \times 10^{-3}\,\mathrm{K}^{-1}$ to $10^{-2}\,\mathrm{K}^{-1}$. 145

List of Tables

5.1 Material properties, and initial and boundary conditions for validation example 1. 44
5.2 Reference and computed values for T_0 concerning the validation example 1. 46
5.3 Material properties, and initial and boundary conditions for validation example 2. 49
5.4 Reference and computed values for T_0 concerning the validation example 2. 50
5.5 Material properties, and initial and boundary conditions for validation example 3. 54
5.6 Reference and computed values for T_0 concerning the validation example 3 in two-dimensions. 55
5.7 Reference and computed values for T_0 concerning the validation example 3 in three-dimensions. 55

6.1 Summary of the comparison between the FFT-Galerkin method. 69

7.1 Summary of the comparison between method for the solution methods of non-linear systems of equations. n here denotes the number of unknowns and m denotes depending on the context the number of previous iterates considered, the number of fixed point evaluations or the size of the Krylov subspace. 102
7.2 Summary of the update formulas for the solution methods of non-linear systems of equations. 104

8.1 Material properties and initial and boundary conditions for the problem concerning the quasi-static finite strain thermo-elastic expansion of an infinitely long thick-walled cylinder. 106
8.2 Total CPU time in seconds and total number of residual evaluations as a function of the thermal expansion coefficient for the best performing implicit methods in each class considered in the solution of the quasi-static expansion of a thermoelastic thick-walled cylinder with $\alpha_T = 5 \times 10^{-5}\,\mathrm{K}^{-1}$, $10 \times 10^{-5}\,\mathrm{K}^{-1}$, and $15 \times 10^{-5}\,\mathrm{K}^{-1}$, and $\dot{u}_0 = 0.5\,\mathrm{mms}^{-1}$. 123
8.3 Material properties and initial and boundary conditions for the problem concerning the quasi-static finite strain thermo-elastic expansion of an infinitely long thick-walled cylinder. 129

8.4 Total CPU time in seconds and the total number of residual evaluations as a function of the thermal softening parameters $w_0 = w_h$ for the best performing implicit methods in each class considered in the solution of the necking of a circular bar with $w_0 = w_h = 2 \times 10^{-3}\,\mathrm{K}^{-1}$, $4.89 \times 10^{-3}\,\mathrm{K}^{-1}$, and $10^{-2}\,\mathrm{K}^{-1}$. 144

Page intentionally left blank.

Chapter 1

Continuum Mechanics and Finite Element Method

This chapter deals with the concepts needed to described the behavior of a solid undergoing large deformation, as well as, the conservation principles that ensure its mechanical equilibrium. It also presents a succint overview of the Finite Element Method as a tool to solve mechanical initial value equilibrium problem. These topics are broadly covered in the literature and here the approach used follows **?**,

1.1 Kinematics of Deformation

1.1.1 Motion

Let a deformable body $\mathscr{B}$ occupy an open region Ω_0 of the tridimensional Euclidean space $\mathscr{E}$ with a regular boundary $\partial\Omega_0$ in its reference configuration. Its motion, depicted in Figure **??**, is defined by a smooth one-to-one function

$$\boldsymbol{\varphi}\colon \Omega \times \mathscr{R} \to \mathscr{E}, \tag{1.1}$$

mapping each material particle of coordinates $\boldsymbol{X}$ in the reference configuration to its position $\boldsymbol{x}$ in the deformed configuration, for a given instant of time t, as

$$\boldsymbol{x} = \boldsymbol{\varphi}(\boldsymbol{X}, t) = \boldsymbol{\varphi}_t(\boldsymbol{X}). \tag{1.2}$$

Thus, the displacement field is defined as

$$\boldsymbol{u}(\boldsymbol{X}, t) = \boldsymbol{\varphi}(\boldsymbol{X}, t) - \boldsymbol{X}, \tag{1.3}$$

and, since the function that defines the motion is one-to-one, the reference configuration can be recovered as

$$\boldsymbol{X} = \boldsymbol{\varphi}^{-1}(\boldsymbol{x}, t) = \boldsymbol{x} - \boldsymbol{u}(\boldsymbol{\varphi}^{-1}(\boldsymbol{x}, t), t), \tag{1.4}$$

where $\boldsymbol{\varphi}^{-1}$ is the reference mapping function.

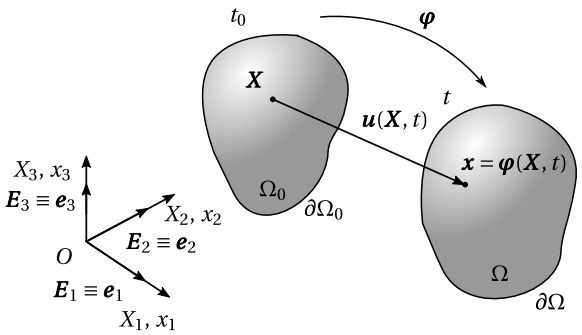


Figure 1.1: Motion

1.1.2 Material and spatial descriptions

Dealing with finite deformations, the behavior of the body under analysis can be described with respect to the reference configuration, using the so-called material or Lagrangian description, or to the deformed configuration, using the so-called spatial or Eulerian description.

In the Lagrangian description any field, be it scalar, vectorial or tensorial defined over the body is expressed as a function of the reference configuration, $\boldsymbol{X} \in \Omega_0$. On the other hand, the Eulerian descritpion of same field is done using the deformed configuration, $\boldsymbol{x} \in \Omega$.

As such let $\alpha(\boldsymbol{x},t)$ be a spatial field and $\beta(\boldsymbol{X},t)$ a material field. Their material α_m and spatial β_s desctiptions are given by

$$\alpha_m(\boldsymbol{X},t) = \alpha(\boldsymbol{\varphi}(\boldsymbol{X},t),t), \tag{1.5}$$

$$\beta_s(\boldsymbol{x},t) = \beta(\boldsymbol{\varphi}^{-1}(\boldsymbol{x},t),t), \tag{1.6}$$

noting that any field associated with a motion of $\mathscr{B}$ can be expressed as a function of time and the point's position in the reference or deformed configuration.

The same distinction between material and spatial descriptions applies to operators such as the divergence and the gradient. The spatial and material gradients, ∇ and ∇_0, respectively, are defined as

$$\nabla\alpha = \frac{\partial}{\partial \boldsymbol{x}}\alpha(\boldsymbol{x},t), \quad \nabla_0\beta = \frac{\partial}{\partial \boldsymbol{X}}\beta(\boldsymbol{X},t), \tag{1.7}$$

where the derivatives are taken with respect to the spatial and reference configuration accordingly.

1.1.3 Deformation gradient

The deformation gradient, a second order tensor denoted by $\boldsymbol{F}$, is defined as

$$\boldsymbol{F}(\boldsymbol{X},t) \equiv \nabla_0\boldsymbol{\varphi}(\boldsymbol{X},t) = \frac{\partial \boldsymbol{x}}{\partial \boldsymbol{X}}, \tag{1.8}$$

or, taking into account that

$$\boldsymbol{x} = \boldsymbol{X} + \boldsymbol{u}(\boldsymbol{X}, t), \tag{1.9}$$

it can be expressed as

$$\boldsymbol{F}(\boldsymbol{X}, t) = \boldsymbol{I} + \nabla_0 \boldsymbol{u}. \tag{1.10}$$

The deformation gradient relates the relative position between two neighboring material particles before and after deformation. To see this let $\boldsymbol{X}$ be the coordinates of some material particle in the reference configuration and $\boldsymbol{X} + d\boldsymbol{X}$ the coordinates of some material particle in its neighborhood, their corresponding coordinates in the deformed configuration are given by

$$\boldsymbol{X} = \boldsymbol{x} - \boldsymbol{u}(\boldsymbol{X}, t), \tag{1.11}$$

$$\boldsymbol{X} + d\boldsymbol{X} = \boldsymbol{x} + d\boldsymbol{x} - \boldsymbol{u}(\boldsymbol{X} + d\boldsymbol{X}, t). \tag{1.12}$$

Subtracting Equation (1.11) to Equation (1.12), it is found that

$$d\boldsymbol{X} = d\boldsymbol{x} + \boldsymbol{u}(\boldsymbol{X}, t) - \boldsymbol{u}(\boldsymbol{X} + d\boldsymbol{X}, t) \tag{1.13}$$

$$= (\boldsymbol{I} + \nabla_0 \boldsymbol{u}(\boldsymbol{X}, t))\, d\boldsymbol{x} \tag{1.14}$$

$$= \boldsymbol{F}\, d\boldsymbol{x}. \tag{1.15}$$

Due to this relation, it can be shown that the determinat of the deformation gradient has a physical meaning. It is the local unit volume change, that is,

$$J \equiv \det \boldsymbol{F} = \frac{dv}{dv_0}, \tag{1.16}$$

where dv_0 is an infinitesimal volume of the body in its reference configuration and dv the infinitesimal volume after deformation.

1.1.3.1 Isochoric/Volumetric decomposition

Any deformation can be locally decomposed in volumetric and isochoric (or distortional) components. From Equation (1.16) it can be gathered that an isochoric deformation is characterized by $J = 1$. As such, the deformation gradient can be decomposed as

$$\boldsymbol{F} = \boldsymbol{F}_{\text{iso}} \boldsymbol{F}_{\text{vol}} = \boldsymbol{F}_{\text{vol}} \boldsymbol{F}_{\text{iso}}, \tag{1.17}$$

where the isochoric and columetric components are defined by

$$\boldsymbol{F}_{\text{iso}} = (\det \boldsymbol{F})^{-\frac{1}{3}}, \quad \boldsymbol{F}_{\text{vol}} = (\det \boldsymbol{F})^{\frac{1}{3}} \boldsymbol{I}. \tag{1.18}$$

1.1.3.2 Polar decomposition

The deformation gradient can also be decomposed in rotation and stretch components, the so-called polar decomposition, defined as

$$\boldsymbol{F} = \boldsymbol{R}\boldsymbol{U} = \boldsymbol{V}\boldsymbol{R}, \tag{1.19}$$

where $\boldsymbol{R}$ is the proper orthogonal rotation tensor and $\boldsymbol{U}$ and $\boldsymbol{V}$ are the symmetric positive right and left stretch tensors, respectively.

Equation (1.19) has a physical interpretation with the right polar decomposition ($\boldsymbol{F} = \boldsymbol{RU}$) corresponding to a stretch mapping followed by a rotation, and the left polar decomposition ($\boldsymbol{F} = \boldsymbol{VR}$) corresponding to a rotation followed by a stretch mapping. The right $\boldsymbol{U}$ and left $\boldsymbol{V}$ stretch tensors are related through the rotation matrix $\boldsymbol{R}$ as

$$\boldsymbol{V} = \boldsymbol{RUR}^T, \tag{1.20}$$

and can be obtained from deformation gradient by

$$\boldsymbol{C} \equiv \boldsymbol{U}^2 = \boldsymbol{F}^T\boldsymbol{F}, \quad \boldsymbol{B} \equiv \boldsymbol{V}^2 = \boldsymbol{FF}^T., \tag{1.21}$$

where $\boldsymbol{C}$ and $\boldsymbol{B}$ are the right and left Cauchy-Green strain tensors.

Since $\boldsymbol{U}$ and $\boldsymbol{V}$ are symmetric tensors, they admit the spectral decomposition

$$\boldsymbol{U} = \sum_{i=1}^{3} \lambda_i \boldsymbol{E}_i^* \otimes \boldsymbol{E}_i^*, \quad \boldsymbol{V} = \sum_{i=1}^{3} \lambda_i \boldsymbol{e}_i^* \otimes \boldsymbol{e}_i^*, \tag{1.22}$$

where λ_i, $i = 1,2,3$, are the eigenvalues of both $\boldsymbol{U}$ and $\boldsymbol{V}$ and $\boldsymbol{E}_i^*$ and $\boldsymbol{e}_i^*$ are the respective eigenvectors.

The eigenvectors of left $\boldsymbol{V}$ and right $\boldsymbol{U}$ stretch tensors are related through

$$\boldsymbol{e}_i^* = \boldsymbol{RE}_i^*. \tag{1.23}$$

forming two orthogonal bases. These vectors define the Lagrangian and Eulerian principal directions, respectively, allowing for the expression of the local stretching from a material particle, associated with any deformation, as a superposition of stretches along the three mutual orthogonal directions.,

1.2 Strain tensors

In Continuum Mechanics there are two main families of strain tensors derived from the deformation gradient and used to describe the body deformation. The Lagrange family strain tensors are defined as

$$\boldsymbol{E}^{(m)} = \begin{cases} \dfrac{1}{m}(\boldsymbol{U}^m - \boldsymbol{I}), & m \neq 0, \\ \ln(\boldsymbol{U}), & m = 0, \end{cases} \tag{1.24}$$

where m is a real number, and likewise, the Euler family strain tensors are defined as

$$\boldsymbol{\varepsilon}^{(m)} = \begin{cases} \dfrac{1}{m}(\boldsymbol{V}^m - \boldsymbol{I}), & m \neq 0, \\ \ln(\boldsymbol{V}), & m = 0, \end{cases} \tag{1.25}$$

where m is also real number.

In particular, choosing $m = 0$, one obtains the so-called material and spatial logarithmic strain tensors

$$\boldsymbol{E}^{(0)} \equiv \ln[\boldsymbol{U}] = \sum_{i=1}^{3} \ln \lambda_i \boldsymbol{E}_i^* \otimes \boldsymbol{E}_i^*, \tag{1.26}$$

$$\boldsymbol{e}^{(0)} \equiv \ln[\boldsymbol{V}] = \sum_{i=1}^{3} \ln \lambda_i \boldsymbol{e}_i^* \otimes \boldsymbol{e}_i^*. \tag{1.27}$$

1.3 Forces and stress measures

The deformation of a body is intrinsically related to the forces acting on it. These forces can be divided in two classes, from a purely mechanical point of view: volume (or body) forces, proportional to the mass contained in a volume element, as such measured in force per unit volume, and surface forces, acting on the surface of a volume element, measured as force per unit area. Related to the latter is the concept of stress, that can be described mathematically by second order tensors with different definitions.

1.3.0.1 Cauchy stress tensor

According to Cauchy's theorem the relation between the so-called Cauchy stress vector, $\boldsymbol{t}(\boldsymbol{x},\boldsymbol{n})$, and the unitary outward vector normal to the deformed surface under analysis, $\boldsymbol{n}$, is linear and given by

$$\boldsymbol{t}(\boldsymbol{x},\boldsymbol{n}) \equiv \boldsymbol{\sigma}(\boldsymbol{x})\boldsymbol{n}, \tag{1.28}$$

where $\boldsymbol{\sigma}$ is the second order Cauchy stress tensor.

The Cauchy stress vector is naturally associated with the deformed configured and thus, expressed in a spatial description and measured in force per unit deformed area. It must also be noted that as a consequence of the balance of angular momentum, the Cauchy stress tensor is symmetric.

1.3.0.2 First Piola-Kirchhoff stress tensor

The First Piola-Kirchhoff stress tensor, $\boldsymbol{P}$, can be regarded as the material counterpart of the Cauchy stress tensor, as it establishes a linear dependence between the stress vector $\boldsymbol{t}_0(\boldsymbol{X},\boldsymbol{m})$, measured in force per unit reference area, and the unitary outward vector normal to the undeformed surface under analysis, $\boldsymbol{m}$,

$$\boldsymbol{t}_0 = \boldsymbol{P}\boldsymbol{m}, \tag{1.29}$$

which must related to the Cauchy stress vector by

$$\boldsymbol{t}_0 = \frac{\mathrm{d}a}{\mathrm{d}a_0}\boldsymbol{t} = \frac{\mathrm{d}a}{\mathrm{d}a_0}\boldsymbol{\sigma}\boldsymbol{n}, \tag{1.30}$$

where $\mathrm{d}a$ is the infinitesimal deformed area normal to the unitary vector $\boldsymbol{n}$ and $\mathrm{d}a_0$ the corresponding undeformed are normal to $\boldsymbol{m}$. It can be shown that the relation between da and da_0 is

$$\frac{da}{da_0}\boldsymbol{n} = J\boldsymbol{F}^{-T}\boldsymbol{m}, \tag{1.31}$$

and substituting on the equation above motivates the following definition

$$\boldsymbol{P} \equiv J\boldsymbol{\sigma}\boldsymbol{F}^{-T}, \tag{1.32}$$

where J is the determinant of the deformation gradient $\boldsymbol{F}$ and $\boldsymbol{\sigma}$ is the Cauchy stress tensor. From Equation (1.32), one gathers that, in general, the First Piola-Kirchhoff stress tensor is not symmetric.

1.3.0.3 Kirchhoff stress tensor

The Kirchhoff stress tensor, $\boldsymbol{\tau}$, is a widely used symmetric tensor, defined as

$$\boldsymbol{\tau} \equiv J\boldsymbol{\sigma}. \tag{1.33}$$

1.3.0.4 Deviatoric/Hydrostatic decomposition

The Cauchy stress tensor, $\boldsymbol{\sigma}$, can split as

$$\boldsymbol{\sigma} = \boldsymbol{\sigma}_d - p\boldsymbol{I}, \tag{1.34}$$

where p is the hydrostatic pressure defined as

$$p \equiv -\frac{1}{3}\text{tr}\,[\boldsymbol{\sigma}], \tag{1.35}$$

and $\boldsymbol{\sigma}_d$ is the deviatoric stress defined as

$$\sigma_d \equiv \boldsymbol{\sigma} - p\boldsymbol{I}. \tag{1.36}$$

1.4 Heat

Heat flowing inside a body, entering or leaving it, often leads to temperature changes. In Continuum Mechanics, heat is measured in power per unit surface.

1.4.0.1 Heat flux vector

According to Cauchy's theorem the relation between the heat flux across a surface, $h(\boldsymbol{x},\boldsymbol{n})$, and the unitary outward normal to the deformed surface under analysis, $\boldsymbol{n}$, is linear and given by

$$h(\boldsymbol{x},\boldsymbol{n}) = -\boldsymbol{q}(\boldsymbol{x})\cdot\boldsymbol{n}. \tag{1.37}$$

where $\boldsymbol{q}$ is the heat flux vector.

1.5 Fundamental conservation principles

In Continuum Mechanics. there is a set of conservation principles and thermodynamic laws, that irrespective of the quantities used to describe the mechanical behavior of a body undergoing large deformations must always be satisfied.

1.5.1 Principle of mass conservation

The principle of mass conservation can be stated as

$$\dot{\rho} + \rho \operatorname{div} \dot{\boldsymbol{u}} = 0, \tag{1.38}$$

where ρ is the material density measured in mass per unit deformed volume.

1.5.2 Principle of linear momentum conservation

The principle of linear momentum conservation can be stated in both material and spatial description. In a spatial description it reads

$$\begin{cases} \operatorname{div} \boldsymbol{\sigma} + \boldsymbol{b} = \rho \ddot{\boldsymbol{u}}, & \forall \boldsymbol{x} \in \Omega, \\ \\ \boldsymbol{t} = \boldsymbol{\sigma} \boldsymbol{n}, & \forall \boldsymbol{x} \in \partial\Omega, \end{cases} \tag{1.39}$$

where $\boldsymbol{b}$ is the body forces field measured as per unit deformed volume.

One can also write the principle of linear momentum conservation in material coordinates, as

$$\begin{cases} \operatorname{div}_0 \boldsymbol{P} + \boldsymbol{b}_0 = \rho_0 \ddot{\boldsymbol{u}}, & \forall \boldsymbol{X} \in \Omega_0, \\ \\ \boldsymbol{t}_0 = \boldsymbol{P} \boldsymbol{m}, & \forall \boldsymbol{X} \in \partial\Omega_0, \end{cases} \tag{1.40}$$

where $\boldsymbol{b}_0$ is the body forces field, measured in force per unit undeformed volume, and ρ_0 is the material density, measured in mass per unit undeformed volume. Both these quantities can be found from their spatial counterparts as

$$\boldsymbol{b}_0 = J\boldsymbol{b}, \quad \rho_0 = J\rho. \tag{1.41}$$

Take notice of the abuse of language regarding functions defined on the referece configuration Ω_0 and on the deformed configuration Ω. The same symbol, f, is used for a function f defined on Ω and the function $f \circ \boldsymbol{\varphi}$ defined on Ω_0.

Equations (1.39) and (1.40) are the so-called strong, point-wise or local equilibrium equations, as they enforce the mechanical equilibrium at every material particle of the body.

1.5.3 First principle of thermodynamics

Let e be the internal energy per unit mass, r the heat supply per unit mass and $\boldsymbol{q}$ the heat flux, then the first principle of thermodynamics pertaining to the balance of energy can be written in the spatial description as

$$\begin{cases} \rho \dot{e} = \boldsymbol{\sigma} : \boldsymbol{D} + \rho r - \operatorname{div} \boldsymbol{q}, & \forall \boldsymbol{x} \in \Omega, \\ \boldsymbol{t} = \boldsymbol{\sigma} \boldsymbol{n}, & \forall \boldsymbol{x} \in \partial\Omega, \\ h = \boldsymbol{q} \cdot \boldsymbol{n}, & \forall \boldsymbol{x} \in \partial\Omega. \end{cases} \tag{1.42}$$

The second order tensor $\boldsymbol{D}$ denotes a strain rate measure, such that the double contraction $\boldsymbol{\sigma}:\boldsymbol{D}$ represents the stress power per unit volume in the deformed configuration of body. In material coordinates, it reads

$$\begin{cases} \rho_0 \dot{e} = \boldsymbol{P}:\dot{\boldsymbol{F}} + \rho_0 r - \operatorname{div}_0 \boldsymbol{q}_0, & \forall \boldsymbol{X} \in \Omega_0, \\ \boldsymbol{t}_0 = \boldsymbol{P}\boldsymbol{m}, & \forall \boldsymbol{X} \in \partial\Omega_0, \\ h_0 = \boldsymbol{q}_0 \cdot \boldsymbol{m}, & \forall \boldsymbol{X} \in \partial\Omega_0, \end{cases} \tag{1.43}$$

where $\boldsymbol{q}_0$ is the Piola transformation of $\boldsymbol{q}$, i.e.,

$$\boldsymbol{q}_0 = J\boldsymbol{F}^{-T}\boldsymbol{q}, \tag{1.44}$$

and

$$h_0 = Jh. \tag{1.45}$$

1.5.4 Second principle of thermodynamics

The local entropy balance can written as

$$\rho \dot{s} = -\operatorname{div}\left[\frac{\boldsymbol{q}}{\theta}\right] + \frac{\rho r}{\theta} + \hat{s}, \tag{1.46}$$

where $\hat{s}$ is the entropy production. The second principle of thermodynamics postulates that the changes in the entropy in the universe can never be negative, which is mathematically expressed by

$$\hat{s} \geq 0, \tag{1.47}$$

yielding

$$\rho \dot{s} + \operatorname{div}\left[\frac{\boldsymbol{q}}{\theta}\right] - \frac{\rho r}{\theta} \geq 0, \tag{1.48}$$

where θ and s are the temperature and specific entropy fields respectively. In a material description, it reads

$$\rho_0 \dot{s} + \operatorname{div}_0\left[\frac{\boldsymbol{q}_0}{\theta}\right] - \frac{\rho_0 r}{\theta} \geq 0. \tag{1.49}$$

1.5.5 Clausius-Duhem inequality

Combining the first and second thermodynamic principles yields

$$\rho \dot{s} + \operatorname{div}\left[\frac{\boldsymbol{q}}{\theta}\right] - \frac{1}{\theta}(\rho \dot{e} - \boldsymbol{\sigma}:\boldsymbol{D} + \operatorname{div}\boldsymbol{q}) \geq 0, \tag{1.50}$$

From the definition of the specific Helmholtz free energy

$$\psi \equiv e - \theta s, \tag{1.51}$$

and defining the temperature field gradient as $\boldsymbol{g} = \nabla\theta$, it is possible to establish the so-called Clausius-Duhem inequality in the spatial description as

$$\underbrace{\boldsymbol{\sigma} : \boldsymbol{D} - \rho\left(\dot{\psi} + s\dot{\theta}\right)}_{\mathcal{D}_{\text{int}}} \underbrace{- \frac{1}{\theta}\boldsymbol{q}\cdot\boldsymbol{g}}_{\mathcal{D}_{\text{cond}}} \geq 0, \tag{1.52}$$

where the identity

$$\operatorname{div}\left[\frac{\boldsymbol{q}}{\theta}\right] = \frac{1}{\theta}\operatorname{div}\boldsymbol{q} - \frac{1}{\theta^2}\boldsymbol{q}\cdot\nabla\theta. \tag{1.53}$$

is used.

From a physical point of view, the Clausius-Duhem inequality states that the energy dissipation per unit deformed volume is always non-negative. Moreover the terms in the inequalitiy can be splited into the internal dissipation $\mathcal{D}_{\text{int}}$ and the dissipation due to heat conduction $\mathcal{D}_{\text{cond}}$. From

$$\hat{s} = \boldsymbol{\sigma} : \boldsymbol{D} - \rho\left(\dot{\psi} + s\dot{\theta}\right) - \frac{1}{\theta}\boldsymbol{q}\cdot\boldsymbol{g}, \tag{1.54}$$

assuming that the process leads to an uniform temperature distribution, yields for the internal dissipation $\mathcal{D}_{\text{int}}$,

$$\mathcal{D}_{\text{int}} = \hat{s}|_{\theta\ \text{uniform}} = \boldsymbol{\sigma} : \boldsymbol{D} - \rho\left(\dot{\psi} + s\dot{\theta}\right), \tag{1.55}$$

since conduction is excluded. If on the other hand, only conduction effects are retained, the dissipation due to conduction, $\mathcal{D}_{\text{cond}}$, is obtained as

$$\mathcal{D}_{\text{cond}} = -\frac{1}{\theta}\boldsymbol{q}\cdot\boldsymbol{g}. \tag{1.56}$$

Equation (1.52) can also be written as

$$\boldsymbol{P} : \dot{\boldsymbol{F}} - \rho_0(\dot{\psi} + s\dot{\theta}) - \frac{1}{\theta}\boldsymbol{q}_0\cdot\boldsymbol{g}_0 \geq 0, \tag{1.57}$$

where $\boldsymbol{g}_0 = \nabla_0\theta$, aplying the Piola transformation, and as

$$\boldsymbol{\tau} : \boldsymbol{D} - \rho_0\left(\dot{\psi} + s\dot{\theta}\right) - \frac{J}{\theta}\boldsymbol{q}\cdot\boldsymbol{g} \geq 0, \tag{1.58}$$

multiplying it by J and attending to the definition of the Kirchhoff stress tensor, where the left hand side represents now the energy dissipation per unit reference volume.

1.6 Thermomechanical constitutive initial value problem

In Continuum Mechanics, a constitutive model is a set of equations, also called constitutive equations, establishing the stress-strain relation for some material. Before going further, it is important to define a thermokinetic process of a body $\mathscr{B}$ as

$$\text{thermokinetic process:} \quad \{\boldsymbol{\varphi}(\boldsymbol{X}, t), \theta(\boldsymbol{X}, t)\}, \tag{1.59}$$

and a calordynamic process of $\mathscr{B}$ as

$$\text{calorodynamic process:} \quad \{\boldsymbol{\sigma}(\boldsymbol{X},t), e(\boldsymbol{X},t), s(\boldsymbol{X},t), r(\boldsymbol{X},t), \boldsymbol{b}(\boldsymbol{X},t), \boldsymbol{q}(\boldsymbol{X},t)\}, \quad (1.60)$$

which satisfies the fundamental conservation principles previously introduced.

It is also important to note that any constitutive model must satisfy a set of constitutive axioms, explained in detail in **?**. As these are too general to be used directly in practice, a particular case of the general history functional-based constitutive theory based on the thermodynamics with internal variables approach is used.

1.6.1 Thermodynamics with internal variables

The values of $\boldsymbol{\sigma}$, ψ, s and $\boldsymbol{q}$ at a material particle define its thermodynamic state, assuming $\boldsymbol{b}$ follows from the balance of linear momentum and r from the energy balance equation. In thermodynamics with interval variables approach, that thermodynamic state is assumed to be completely defined by the instantaneous values of a finite number of state variables

$$\{\boldsymbol{F}, \theta, \boldsymbol{g}, \boldsymbol{\alpha}\}. \quad (1.61)$$

at a given instant of the calorodynamic process, where

$$\boldsymbol{\alpha} = \{\alpha_k\} \quad (1.62)$$

is a set of internal variables, scalar or tensorial in nature, associated with dissipative mechanisms. As such, the accuracy of the constitutive model depends strongly on the appropriate choice of internal variables, as these contain the relevant information about the thermodynamical history of the material.

Accordingly, the specific Helmholtz free energy is postulated to follow

$$\psi = \psi(\boldsymbol{F}, \theta, \boldsymbol{\alpha}). \quad (1.63)$$

To find the constitutive equations for the stress tensor and the entropy, one can substitute

$$\dot{\psi} = \frac{\partial \psi}{\partial \boldsymbol{F}} : \dot{\boldsymbol{F}} + \frac{\partial \psi}{\partial \theta}\dot{\theta} + \frac{\partial \psi}{\partial \alpha_k}\dot{\alpha}_k, \quad (1.64)$$

found from the chain rule, into the Clausius-Duhem equation, Equation (1.52), obtaining

$$\left(\boldsymbol{\sigma}\boldsymbol{F}^{-T} - \rho\frac{\partial \psi}{\partial \boldsymbol{F}}\right) : \dot{\boldsymbol{F}} - \rho\left(s + \frac{\partial \psi}{\partial \theta}\right)\dot{\theta} - \rho\frac{\partial \psi}{\partial \alpha_k}\dot{\alpha}_k - \frac{1}{\theta}\boldsymbol{q}\cdot\boldsymbol{g} \geq 0, \quad (1.65)$$

where the velocity gradient was adopted to set the work conjugacy as

$$\boldsymbol{\sigma} : \boldsymbol{D} = \boldsymbol{\sigma} : \boldsymbol{L} = \boldsymbol{\sigma} : \dot{\boldsymbol{F}}\boldsymbol{F}^{-1} = \boldsymbol{\sigma}\boldsymbol{F}^{-T} : \boldsymbol{F}. \quad (1.66)$$

Since the Clausius-Duhem inequality must hold for any thermokinetic process and so remain valid for any set $\{\dot{\boldsymbol{F}}(t), \dot{\theta}(t)\}$, the Cauchy stress and entropy constitutive equations must be

$$\boldsymbol{\sigma} = \rho \frac{\partial \psi}{\partial \boldsymbol{F}} \boldsymbol{F}^T, \tag{1.67}$$

$$s = -\frac{\partial \psi}{\partial \theta}. \tag{1.68}$$

It is also possible to write the constitutive equations for the Kirchhoff stress tensor as

$$\boldsymbol{\tau} = J\rho \frac{\partial \psi}{\partial \boldsymbol{F}} \boldsymbol{F}^T, \tag{1.69}$$

multiplying Equation (1.67) by J, and the first Piola-Kirchhoff stress tensor as

$$\boldsymbol{P} = \rho_0 \frac{\partial \psi}{\partial \boldsymbol{F}} \tag{1.70}$$

multiplying Equation (1.65) also by J.

For each internal variable α_k of the set α of internal variables, the conjugate thermodynamical forces are defined to be

$$A_k \equiv \rho_0 \frac{\partial \psi}{\partial \alpha_k}, \tag{1.71}$$

so that the Clausius-Duhem equation can be written in a reduced form as

$$-\boldsymbol{A} * \dot{\boldsymbol{\alpha}} - \frac{J}{\theta} \boldsymbol{q} \cdot \boldsymbol{g} \geq 0, \tag{1.72}$$

where $\boldsymbol{A}$ is the set of conjugate thermodynamical forces and $*$ denotes the appropriate product operation.

To completely define the constitutive model, one still needs to postulate the constitutive equations for the flux variables $\dot{\boldsymbol{\alpha}}$ and $\frac{1}{\theta}\boldsymbol{q}$. These are given by

$$\dot{\boldsymbol{\alpha}} = f(\boldsymbol{F}, \theta, \boldsymbol{g}, \boldsymbol{\alpha}), \tag{1.73}$$

$$\frac{1}{\theta}\boldsymbol{q} = g(\boldsymbol{F}, \theta, \boldsymbol{g}, \boldsymbol{\alpha}). \tag{1.74}$$

A sufficient condition for the previous constitutive functions to satisfy the Clausius-Duhem inequality is the hypothesis of normal dissipativity, whereby one defines the constitutive functions for the flux variables as

$$\dot{\alpha}_k = -\frac{\partial \Xi}{\partial A_k}, \quad \frac{1}{\theta}\boldsymbol{q} = -\frac{\partial \Xi}{\partial \boldsymbol{g}}, \tag{1.75}$$

where the dissipation potential is

$$\Xi = \Xi(\boldsymbol{A}, \boldsymbol{g}; \boldsymbol{F}, \theta, \boldsymbol{\alpha}), \tag{1.76}$$

a convex function with respect to each A_k and $\boldsymbol{g}$, and zero valued at the origin, $\{\boldsymbol{A}, \boldsymbol{g}\} = \{\boldsymbol{0}, \boldsymbol{0}\}$. Note that in the previous definition the state variables appear only as parameters.

Page intentionally left blank.

Chapter 2

Mechanical problem

In the following chapter, the general framework presented in the previous chapter is applied to a purely mechanical analysis, neglecting the thermal terms.

2.0.1 Mechanical constitutive initial value problem

In the purely mechanical case, with all the quantities related to the thermal domain removed, a constitutive model based on internal variables is established by the following set of equations

$$\boldsymbol{P} = \rho_0 \frac{\partial \psi}{\partial \boldsymbol{F}}, \tag{2.1}$$

$$\psi = \psi(\boldsymbol{F}, \boldsymbol{\alpha}), \tag{2.2}$$

$$\dot{\boldsymbol{\alpha}} = f(\boldsymbol{F}, \boldsymbol{\alpha}). \tag{2.3}$$

Thus, the spatial mechanical constitutive initial value problem can be stated as follows

Problem 2.1 | Spatial mechanical constitutive initial value problem.

Given the initial values of the internal variables, $\boldsymbol{\alpha}(t_0)$, and the history of the deformation gradient

$$\boldsymbol{F}(t), \quad t \in [t_0, t_{\text{end}}], \tag{2.4}$$

find the functions for $\boldsymbol{\sigma}(t)$ and $\boldsymbol{\alpha}(t)$ such that the constitutive equations

$$\boldsymbol{\sigma} = \rho \frac{\partial \psi}{\partial \boldsymbol{F}} \boldsymbol{F}^T, \tag{2.5}$$

$$\psi = \psi(\boldsymbol{F}, \boldsymbol{\alpha}), \tag{2.6}$$

$$\dot{\boldsymbol{\alpha}} = f(\boldsymbol{F}, \boldsymbol{\alpha}). \tag{2.7}$$

are satisfied for every $t \in [t_0, t_{\text{end}}]$.

Likewise, in a material description it can be stated as

Problem 2.2 | Material mechanical constitutive intial value problem.

Given the initial values of the internal variables, $\boldsymbol{\alpha}(t_0)$, and the history of the deformation gradient

$$\boldsymbol{F}(t), \quad t \in [t_0, t_{\text{end}}], \tag{2.8}$$

find the functions for $\boldsymbol{P}(t)$ and $\boldsymbol{\alpha}(t)$ such that the constitutive equations

$$\boldsymbol{P} = \rho_0 \frac{\partial \psi}{\partial \boldsymbol{F}}, \tag{2.9}$$

$$\psi = \psi(\boldsymbol{F}, \boldsymbol{\alpha}), \tag{2.10}$$

$$\dot{\boldsymbol{\alpha}} = f(\boldsymbol{F}, \boldsymbol{\alpha}). \tag{2.11}$$

are satisfied for every $t \in [t_0, t_{\text{end}}]$.

2.0.2 Weak equilibrium. The principle of virtual work

The strong equations that enforce the equilirium of a body can be writen using the spatial description as

$$\rho \ddot{\boldsymbol{u}} = \operatorname{div} \boldsymbol{\sigma} + \boldsymbol{b} \quad \text{in } \Omega, \tag{2.12}$$

and the material description as

$$\rho_0 \ddot{\boldsymbol{u}} = \operatorname{div}_0 \boldsymbol{P} + \boldsymbol{b}_0 \quad \text{in } \Omega_0. \tag{2.13}$$

From a practical standpoint, finding the exact solution to the strong equilibrium equations in the context of real engineering problems is most often nearly or entirely impossible. Most numerical methods obtain only approximate solutions to the so-called weak equilibrium equations to circumvent this problem. These result from relaxing the strong equilibrium equations so that the solutions need only satisfy the equilibrium equations in an average sense instead of satisfying them pointwise. This is achieved through an integration over the body volume. The weak equilibrium equations can be found making use of several energetic and weighted residual methods, such as the Virtual Work Principle used here.

Problem 2.3 | Principle of virtual work (spatial version).

The Virtual Work Principle states, in a spatial description, that the body is in equilibrium if and only if the Cauchy stress field satisfies

$$\int_\Omega [\boldsymbol{\sigma} : \nabla \boldsymbol{\eta} - (\boldsymbol{b} - \rho \ddot{\boldsymbol{u}}) \cdot \boldsymbol{\eta}] \, \mathrm{d}v - \int_{\partial\Omega} \boldsymbol{t} \cdot \boldsymbol{\eta} \, \mathrm{d}a = 0, \quad \forall \boldsymbol{\eta} \in \mathscr{V}_u, \tag{2.14}$$

where $\mathscr{V}_u$ is the space of virtual displacement of the body, defined by the space of sufficiently regular arbitrary displacements

$$\boldsymbol{\eta} \colon \Omega \to \mathscr{U} \tag{2.15}$$

where $\mathscr{U}$ is the n-dimension vector associated with $\mathscr{E}$.

The principle of virtual work can be expressed in a completly equivalent way using a material description.

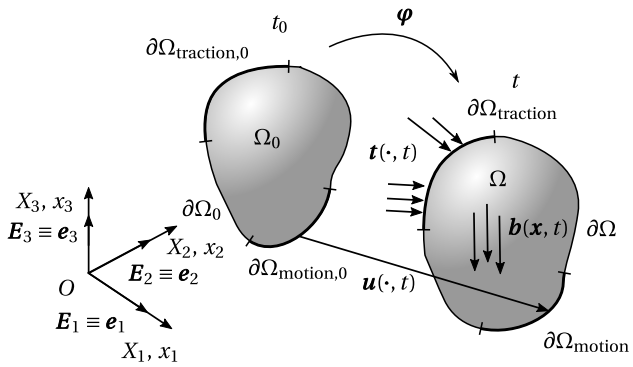


Figure 2.1: Quasi-static mechanical constitutive initial boundary value problem.

Problem 2.4 | Principle of virtual work (material version).

The Virtual Work Principle states, in a material description, that the body is in equilibrium if and only if the First Piola-Kirchhoff stress field satisfies

$$\int_{\Omega_0} [\boldsymbol{P} : \nabla_0 \boldsymbol{\eta} - (\boldsymbol{b}_0 - \rho_0 \ddot{\boldsymbol{u}}) \cdot \boldsymbol{\eta}] \, \mathrm{d}v - \int_{\partial\Omega_0} \boldsymbol{t}_0 \cdot \boldsymbol{\eta} \, \mathrm{d}da = 0, \quad \forall \boldsymbol{\eta} \in \mathscr{V}_{u,0}, \tag{2.16}$$

where $\mathscr{V}_{u,0}$ is the space of virtual displacement of the body, defined by the space of sufficiently regular arbitrary displacements

$$\boldsymbol{\eta} : \Omega_0 \to \mathscr{U}. \tag{2.17}$$

2.0.3 Mechanical constitutive initial boundary value problem

It is now possible to pose the mechanical constitutive initial value problem in its weak form. Assume that a body $\mathscr{B}$ is made from a generic material, characterized by a given constitutive model, whose internal variables are known at the initial time, as presented in Figure **??**. In addition, it is assumed that the interior of the body was subjected to a prescribed history of body forces, $\boldsymbol{b}(\boldsymbol{X},t)$, $t \in [t_0, t_{\text{end}}]$, and to the following boundary conditions:

- **Natural (or Neumann) boundary condition:** The boundary portion $\Omega_{\text{traction},\,0}$ of $\mathscr{B}$ is subjected to a prescribed history of traction forces, $\boldsymbol{t}_{\text{presc}}(\boldsymbol{X},t)$, $\boldsymbol{X} \in \partial\Omega_{\text{traction},0}$, $t \in [t_0, t_{\text{end}}]$,

- **Essential (or Dirichlet) boundary condition:** The boundary portion $\Omega_{\text{motion},\,0}$ of $\mathscr{B}$ is subjected to a prescribed displacement field history, $\boldsymbol{u}_{\text{presc}}(\boldsymbol{X},t)$, such that

$$\boldsymbol{\varphi}(\boldsymbol{X},t) = \boldsymbol{X} + \boldsymbol{u}_{\text{presc}}(\boldsymbol{X},t), \quad \boldsymbol{X} \in \partial\Omega_{\text{motion},\,0}, \quad t \in [t_0, t_{\text{end}}].$$

It is also convenient to define the set of kinematically admissible displacements of $\mathscr{B}$ as the set of all sufficiently regular displacement functions tha satisfy the essential boundary condition (**?**),

$$\mathscr{K}_u \equiv \{\boldsymbol{u} : \Omega_0 \times \mathscr{R} \to \mathscr{U} \mid \boldsymbol{u}(\boldsymbol{X}, t) = \boldsymbol{u}_{\text{presc}}(\boldsymbol{X}, t), \quad \boldsymbol{X} \in \partial\Omega_{\text{motion},0}, \quad t \in [t_0, t_{\text{end}}]\}. \tag{2.18}$$

So the weak form of the quasi-static mechanical constitutive initial boundary value problem can be stated in a spatial description as follows

Problem 2.5 | Spatial mechanical initial BVP.

Find a kinematically admissible displacement function, $\boldsymbol{u} \in \mathscr{K}_u$, such that for every $t \in [t_0, t_{\text{end}}]$, the body $\mathscr{B}$ is in equilibrium as stated by the Virtual Work Principle

$$\int_\Omega [\boldsymbol{\sigma} : \nabla\boldsymbol{\eta} - (\boldsymbol{b} - \rho\ddot{\boldsymbol{u}}) \cdot \boldsymbol{\eta}] \, \mathrm{d}v - \int_{\partial\Omega} \boldsymbol{t} \cdot \boldsymbol{\eta} \, \mathrm{d}a = 0, \quad \forall \boldsymbol{\eta} \in \mathscr{V}_u, \tag{2.19}$$

where the space of virtual displacements at time t is defined by

$$\mathscr{V}_u \equiv \left\{\boldsymbol{\eta} : \Omega \to \mathscr{U} \mid \boldsymbol{\eta} = \boldsymbol{0} \quad \text{in} \quad \boldsymbol{\varphi}(\partial\Omega_{\text{motion},0}, t)\right\}, \tag{2.20}$$

and at each point of $\mathscr{B}$, the Cauchy stress tensor is the solution of spatial mechanical constitutive initial values problem.

and in the material description as

Problem 2.6 | Material mechanical initial BVP.

Find a kinematically admissible displacement function, $\boldsymbol{u} \in \mathscr{K}_u$, such that for every $t \in [t_0, t_{\text{end}}]$, the body $\mathscr{B}$ is in equilibrium as stated by the Virtual Work Principle

$$\int_{\Omega_0} [\boldsymbol{P} : \nabla_0\boldsymbol{\eta} - (\boldsymbol{b}_0 - \rho_0\ddot{\boldsymbol{u}}) \cdot \boldsymbol{\eta}] \, \mathrm{d}v - \int_{\partial\Omega_0} \boldsymbol{t}_0 \cdot \boldsymbol{\eta} \, \mathrm{d}a = 0, \quad \forall \boldsymbol{\eta} \in \mathscr{V}_{u,0}, \tag{2.21}$$

where the space of virtual displacements at time t is defined by

$$\mathscr{V}_{u,0} \equiv \left\{\boldsymbol{\eta} : \Omega_0 \to \mathscr{U} \mid \boldsymbol{\eta} = \boldsymbol{0} \quad \text{in} \quad \partial\Omega_{\text{motion},0}\right\}, \tag{2.22}$$

and at each point of $\mathscr{B}$, the First Piola-Kirchhoff stress tensor is the solution of material mechanical constitutive initial values problem.

2.1 Time descretization

Given a generic path-dependent model, i.e., a model in which the stress state does not depend only on the instantaneous deformation state but also on the deformation history, the solution of the constitutive initial value problem for a given set of initial conditions is usually not known for complex strain paths $\boldsymbol{F}(t)$. Thus, there is a need to use an appropriate numerical algorithm to integrate the rate constitutive equations.

In general, the algorithms for integrating rate constitutive equations are obtained adopting some time (or pseudo-time) discretization and some hypothesis on the deformation path between adjacent time stations. In the present document, an algorithm is adopted based on approximated incremental constitutive functions.

Attending to the mechanical constitutive initial boundary value problem and considering the time increment $[t_n, t_{n+1}]$, this approach is comprised by the following two requirements:

- **Cauchy and First Piola-Kirchhoff stress tensors.** Considering a time increment $[t_n, t_{n+1}]$ and given the set $\boldsymbol{\alpha}_n$ of internal variables at t_n, the deformation gradient $\boldsymbol{F}_{n+1}$ at time t_{n+1} determines the stress $\boldsymbol{\sigma}_{n+1}$ uniquely through

$$\boldsymbol{\sigma}_{n+1} = \hat{\boldsymbol{\sigma}}(\boldsymbol{\alpha}_n, \boldsymbol{F}_{n+1}), \tag{2.23}$$

  where $\hat{\boldsymbol{\sigma}}$ is the incremental constitutive function for the Cauchy stress tensor.

  Similarly, the First Piola-Kirchhoff stress tensor $\boldsymbol{P}_{n+1}$ must be uniquely determined by the prescribed deformation gradient $\boldsymbol{F}_{n+1}$ prescribed at t_{n+1} as

$$\boldsymbol{P}_{n+1} = \hat{\boldsymbol{P}}(\boldsymbol{\alpha}_n, \boldsymbol{F}_{n+1}), \tag{2.24}$$

  where $\hat{\boldsymbol{P}}$ is the incremental constitutive function for the First Piola-Kirchhoff stress tensor.

- **Set of internal variables.** Assuming that the set of internal variables $\boldsymbol{\alpha}_n$ is known at t_n, the set of internal variables must be uniquely determined by the prescribed deformationgradient $\boldsymbol{F}_{n+1}$ prescribed at t_{n+1} as

$$\boldsymbol{\alpha}_{n+1} = \hat{\boldsymbol{\alpha}}(\boldsymbol{\alpha}_n, \boldsymbol{F}_{n+1}), \tag{2.25}$$

  where $\hat{\boldsymbol{\alpha}}$ is the incremental constitutive function for the set of internal variables.

Generally, the numerical constitutive laws are nonlinear and path-independent within one increment. In other words, within each increment, $\boldsymbol{\sigma}_{n+1}$ and $\boldsymbol{\alpha}_{n+1}$, they are functions of $\boldsymbol{F}_{n+1}$ alone with the argument $\boldsymbol{\alpha}_n$ constant within the same time interval.

Making use of the aforementioned time discretization, one can state the weak form of the mechanical constitutive initial boundary value problem in the spatial description as

Problem 2.7 | Spatial incremental mechanical initial BVP.

Given the set of internal variables $\boldsymbol{\alpha}_n$ at t_n, the prescribed body and traction force fields $\boldsymbol{b}_{n+1}$ and $\boldsymbol{t}_{n+1}$ at t_{n+1}, and the prescribed deformating gradient $\boldsymbol{F}_{n+1}$ at t_{n+1}, find the kinematically admissible displacement field $\boldsymbol{u}_{n+1} \in \mathscr{K}_{u,n+1}$ such that the body $\mathscr{B}$ is in equilibrium as stated by the virtual Work Principle

$$\int_{\Omega_{n+1}} [\hat{\boldsymbol{\sigma}}(\boldsymbol{F}_{n+1}, \boldsymbol{\alpha}_n) : \nabla\boldsymbol{\eta} - (\boldsymbol{b}_{n+1} - \rho\ddot{\boldsymbol{u}}_{n+1}) \cdot \boldsymbol{\eta}] \, \mathrm{d}v - \int_{\partial\Omega_{n+1}} \boldsymbol{t}_{n+1} \cdot \boldsymbol{\eta} \, \mathrm{d}a = 0, \quad \forall \boldsymbol{\eta} \in \mathscr{V}_u, \tag{2.26}$$

where the space of kinematically admissible displacement fields $\mathscr{K}_{n+1}$ is defined by

$$\mathscr{K}_{u,n+1} \equiv \{\boldsymbol{u} : \Omega_0 \times \mathscr{R} \to \mathscr{U} \mid \boldsymbol{u}_{n+1}(\boldsymbol{X}) = \boldsymbol{u}_{\mathrm{presc},n+1}(\boldsymbol{X}),\ \boldsymbol{X} \in \partial\Omega_{\mathrm{motion},0}\}. \tag{2.27}$$

and in the material description as

Problem 2.8 | Material incremental mechanical initial BVP.

Given the set of internal variables $\boldsymbol{\alpha}_n$ at t_n, the prescribed body and traction force fields $\boldsymbol{b}_{0,n+1}$ and $\boldsymbol{t}_{0,n+1}$ at t_{n+1}, and the prescribed deformating gradient $\boldsymbol{F}_{n+1}$ at t_{n+1}, find the kinematically admissible displacement field $\boldsymbol{u}_{n+1} \in \mathscr{K}_{u,n+1}$ such that the body $\mathscr{B}$ is in equilibrium as stated by the virtual Work Principle

$$\int_{\Omega_0} [\hat{\boldsymbol{P}}(\boldsymbol{F}_{n+1}, \boldsymbol{\alpha}_n) : \nabla_0 \boldsymbol{\eta} - (\boldsymbol{b}_{0,n+1} - \rho_0 \ddot{\boldsymbol{u}}_{n+1}) \cdot \boldsymbol{\eta}] \, \mathrm{d}v - \int_{\partial\Omega_{0,n+1}} \boldsymbol{t}_{0,n+1} \cdot \boldsymbol{\eta} \, \mathrm{d}a = 0, \quad \forall \boldsymbol{\eta} \in \mathscr{V}_{u,0}, \tag{2.28}$$

where the space of kinematically admissible displacement fields $\mathscr{K}_{n+1}$ is defined by

$$\mathscr{K}_{u,n+1} \equiv \{\boldsymbol{u} : \Omega_0 \times \mathscr{R} \to \mathscr{U} \mid \boldsymbol{u}_{n+1}(\boldsymbol{X}) = \boldsymbol{u}_{\text{presc},n+1}(\boldsymbol{X}),\ \boldsymbol{X} \in \partial\Omega_{\text{motion},0}\}. \tag{2.29}$$

2.2 Finite Element Method

With the incremental weak form of the mechanical constitutive initial boundary value problem now established, an approximated solution can be found using the Finite Element Method.

2.2.1 Finite element concept

The first in the Finite Element method is to discretize the continuum domain Ω in a finite set of n_{elem} mutually exclusive subdomains called finite elements $\Omega^{(e)}$. The discretized domain, ${}^h\Omega$, is therefore an approximation to the continuum domain expressed by

$$\Omega \approx {}^h\Omega \equiv \bigcup_{e=1}^{n_{\text{elem}}} \Omega^{(e)}. \tag{2.30}$$

The spaces of virtual displacements $\mathscr{V}_u$ and $\mathscr{V}_{u,0}$ as well as the space of kinematically admissible displacement fields $\mathscr{K}_u$ are also discretized in the same way, with their discretized forms denoted by ${}^h\mathscr{V}_u$, ${}^h\mathscr{V}_{u,0}$ and ${}^h\mathscr{K}_u$.

2.2.2 Interpolation functions

Let e be a generic finite element with n_{nodes} nodes, where each node i of coordinates $\boldsymbol{x}^i$ is associated with an interpolation function $N_i^{(e)}$. These interpolation functions are often called shape functions and perform the required filed interpolations inside the element domain $\Omega^{(e)}$.

Letting $a(\boldsymbol{x})$ be a generic field defined over $\Omega^{(e)}$, its interpolation at any point $\boldsymbol{x}$ inside the element is defined by the element shape functions as

$$a(\boldsymbol{x}) \approx {}^h a(\boldsymbol{x}) \equiv \sum_{i=1}^{n_{\text{nodes}}} a(\boldsymbol{x}_i) N_i^{(e)}(\boldsymbol{x}). \tag{2.31}$$

If instead $a(\boldsymbol{x})$ is instead a generic field defined over the global domain Ω, the interpolation of $a(\boldsymbol{x})$ at any point $\boldsymbol{x}$ is defined by the global shape functions as

$$a(\boldsymbol{x}) \approx {}^{h}a(\boldsymbol{x}) \equiv \sum_{i=1}^{n_{\text{points}}} a(\boldsymbol{x}_i) N_i^g(\boldsymbol{x}), \tag{2.32}$$

where n_{points} is the total number of nodes of the finite element mesh. The discretized spaces ${}^{h}\mathscr{V}_u$ and ${}^{h}\mathscr{K}_u$ can now be defined as

$${}^{h}\mathscr{K}_u \equiv \left\{ {}^{h}\boldsymbol{u}(\boldsymbol{x}) = \sum_{i=1}^{n_{\text{points}}} \boldsymbol{u}(\boldsymbol{x}_i) N_i^g(\boldsymbol{x}) \mid \boldsymbol{u}(\boldsymbol{x}_i) = \boldsymbol{u}_{\text{presc}}(\boldsymbol{x}_i) \quad \text{if } \boldsymbol{x}_i \in \partial\Omega_{\text{motion},0} \right\}, \tag{2.33}$$

$${}^{h}\mathscr{V}_u \equiv \left\{ {}^{h}\boldsymbol{\eta}(\boldsymbol{x}) = \sum_{i=1}^{n_{\text{points}}} \boldsymbol{\eta}(\boldsymbol{x}_i) N_i^g(\boldsymbol{x}) \mid \boldsymbol{\eta}(\boldsymbol{x}_i) = \boldsymbol{0} \quad \text{if } \boldsymbol{x}_i \in \partial\Omega_{\text{motion},0} \right\} \tag{2.34}$$

Quantities defined on the reference configuration Ω_0 accepted a treatment entirely similar to the one described above, and thus is omitted.

2.2.3 Interpolation matrix and discrete gradient operators

The global shape functions can be conveniently assembled in the so-called global interpolation matrix as

$$\mathbf{N}^g(\boldsymbol{x}) \equiv \left[\text{diag}[N_1^g(\boldsymbol{x})] \; \text{diag}[N_2^g(\boldsymbol{x})] \; \cdots \; \text{diag}[N_{n_{\text{points}}}^g(\boldsymbol{x})] \right], \tag{2.35}$$

where $\text{diag}[N_i^g]$ is a diagonal matrix $n_{\text{dim}} \times n_{\text{dim}}$

$$\text{diag}[N_i^g(\boldsymbol{x})] \equiv \begin{bmatrix} N_i^g & 0 & \cdots & 0 \\ 0 & N_i^g & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & N_i^g \end{bmatrix} \tag{2.36}$$

where n_{dim} is the number of degrees of freedom per node.

Defining the global vector of nodal displacements as

$$\mathbf{u} = \left[u_1^1, \ldots, u_{n_{\text{dim}}}^1, \ldots, u_1^{n_{\text{points}}}, \ldots, u_{n_{\text{dim}}}^{n_{\text{points}}} \right]^T, \tag{2.37}$$

the displacement field $\boldsymbol{u}(\boldsymbol{x})$ defined over the global domain Ω, can be found from Equation (2.32) at any point $\boldsymbol{x}$ as

$${}^{h}\boldsymbol{u}(\boldsymbol{x}) \equiv \mathbf{N}^g(\boldsymbol{x})\mathbf{u}, \quad {}^{h}\boldsymbol{u} \in {}^{h}\mathscr{K}_u. \tag{2.38}$$

2.2.4 Spatial discretization

Applying the aforementioned finite element discretization to the incremental mechanical constitutive initial boundary value problem, we can then write in the spatial description

$$\int_{{}^h\Omega}\left[\hat{\boldsymbol{\sigma}}^T\mathbf{B}^g\boldsymbol{\eta}-(\mathbf{b}_{n+1}-\rho\ddot{\mathbf{u}}_{n+1})\cdot\mathbf{N}^g\boldsymbol{\eta}\right]\mathrm{d}v-\int_{\partial^h\Omega_{\text{traction}}}\mathbf{t}_{n+1}\cdot\mathbf{N}^g\boldsymbol{\eta}\,\mathrm{d}a=0,\quad\forall\boldsymbol{\eta}\in{}^h\mathscr{V}_u, \tag{2.39}$$

where $\mathbf{B}^g$ is the discrete symmetric global gradient operator, defined for a 2D problem in cartesian coordinates as

$$\mathbf{B}^g\equiv\begin{bmatrix}\frac{\partial N_1^g}{\partial x} & 0 & \frac{\partial N_2^g}{\partial x} & 0 & \cdots & \frac{\partial N_{n_{\text{points}}}^g}{\partial x} & 0\\ 0 & \frac{\partial N_1^g}{\partial y} & 0 & \frac{\partial N_2^g}{\partial y} & \cdots & 0 & \frac{\partial N_{n_{\text{points}}}^g}{\partial y}\\ \frac{\partial N_1^g}{\partial y} & \frac{\partial N_1^g}{\partial x} & \frac{\partial N_2^g}{\partial y} & \frac{\partial N_2^g}{\partial x} & \cdots & \frac{\partial N_{n_{\text{points}}}^g}{\partial y} & \frac{\partial N_{n_{\text{points}}}^g}{\partial x}\end{bmatrix}. \tag{2.40}$$

Equation (2.39) can be rewritten as

$$\left\{\int_{{}^h\Omega}\left[\mathbf{B}^{gT}\hat{\boldsymbol{\sigma}}(\boldsymbol{\alpha}_n,\boldsymbol{F}_{n+1})-\mathbf{N}^{gT}\mathbf{b}_{n+1}+\mathbf{N}^{gT}\rho\ddot{\mathbf{u}}_{n+1}\right]\mathrm{d}v-\int_{\partial^h\Omega_{\text{traction}}}\mathbf{N}^{gT}\mathbf{t}_{n+1}\,\mathrm{d}a\right\}^T\boldsymbol{\eta}=0,\quad\forall\boldsymbol{\eta}\in{}^h\mathscr{V}_u, \tag{2.41}$$

and, since it must be satisfied for any $\boldsymbol{\eta}\in{}^h\mathscr{V}_u$, the incremental quasi-static discretized mechanical constitutive initial boundary value problem can thus be stated in the spatial description as

Problem 2.9 | Spatial incremental discretized mechanical initial BVP.

Given the set of internal variables $\boldsymbol{\alpha}_n$ at t_n, the prescribed body and traction force fields $\boldsymbol{b}_{n+1}$ and $\boldsymbol{t}_{n+1}$, and the prescribed deformation gradient $\boldsymbol{F}_{n+1}$ at t_{n+1}, find the kinematically admissible nodal displacement field $\boldsymbol{u}_{n+1}\in{}^h\mathscr{K}_{u,n+1}$ such that the body $\mathscr{B}$ is in equilibrium as stated by the Virtual Work Principle

$$\mathbf{M}\ddot{\mathbf{u}}_{n+1}+\mathbf{f}^{\text{int}}(\mathbf{u}_{n+1})-\mathbf{f}_{n+1}^{\text{ext}}=\mathbf{0}, \tag{2.42}$$

where $\mathbf{f}^{\text{int}}$ e $\mathbf{f}_{n+1}^{\text{ext}}$ are the global vectors of internal and external forces defined as

$$\mathbf{f}^{\text{int}}\equiv\int_{{}^h\Omega}\mathbf{B}^{gT}\hat{\boldsymbol{\sigma}}(\boldsymbol{F}_{n+1},\boldsymbol{\alpha}_n)\,\mathrm{d}v, \tag{2.43}$$

$$\mathbf{f}_{n+1}^{\text{ext}}\equiv\int_{{}^h\Omega}\mathbf{N}^{gT}\mathbf{b}_{n+1}\,\mathrm{d}v+\int_{\partial^h\Omega_{\text{traction}}}\mathbf{N}^{gT}\mathbf{t}_{n+1}\,\mathrm{d}a, \tag{2.44}$$

and $\mathbf{M}$ is the mass matrix defined as

$$\mathbf{M}=\int_{{}^h\Omega}\rho\mathbf{N}^{gT}\mathbf{N}^g\,\mathrm{d}v. \tag{2.45}$$

In a material description, Equation (2.41) is written as

$$\left\{\int_{{}^h\Omega_0}\left[\mathbf{G}^{g\,T}\hat{\boldsymbol{P}}(\boldsymbol{\alpha}_n,\boldsymbol{F}_{n+1})-\mathbf{N}^{g\,T}\mathbf{b}_{0,n+1}+\mathbf{N}^{g\,T}\rho_0\ddot{\mathbf{u}}_{n+1}\right]\mathrm{d}v\right.$$
$$\left.-\int_{\partial^h\Omega_{\text{traction},0}}\mathbf{N}^{g\,T}\mathbf{t}_{0,n+1}\,\mathrm{d}a\right\}^T\boldsymbol{\eta}=0,\quad\forall\boldsymbol{\eta}\in{}^h\mathscr{V}_{u,0},\tag{2.46}$$

where $\mathbf{G}^g$ is the discrete global gradient operator, defined for a 2D problem in cartesian coordinates as

$$\mathbf{G}^g\equiv\begin{bmatrix}\frac{\partial N_1^g}{\partial x} & 0 & \frac{\partial N_2^g}{\partial x} & 0 & \cdots & \frac{\partial N_{n_{\text{points}}}^g}{\partial x} & 0\\ 0 & \frac{\partial N_1^g}{\partial x} & 0 & \frac{\partial N_2^g}{\partial x} & 0 & \cdots & \frac{\partial N_{n_{\text{points}}}^g}{\partial x}\\ \frac{\partial N_1^g}{\partial y} & 0 & \frac{\partial N_2^g}{\partial y} & \cdots & 0 & \frac{\partial N_{n_{\text{points}}}^g}{\partial y} & 0\\ 0 & \frac{\partial N_1^g}{\partial y} & 0 & \frac{\partial N_2^g}{\partial y} & \cdots & 0 & \frac{\partial N_{n_{\text{points}}}^g}{\partial y}\end{bmatrix},\tag{2.47}$$

and, as for the spatial description, it must be satisfied for any $\boldsymbol{\eta}\in{}^h\mathscr{V}_{u,0}$, the incremental quasi-static discretized mechanical constitutive initial boundary value problem can thus be stated in the material description as

Problem 2.10 | Material incremental discretized mechanical initial BVP.

Given the set of internal variables $\boldsymbol{\alpha}_n$ at t_n, the prescribed body and traction force fields $\boldsymbol{b}_{0,n+1}$ and $\boldsymbol{t}_{0,n+1}$, and the prescribed deformation gradient $\boldsymbol{F}_{n+1}$ at t_{n+1}, find the kinematically admissible nodal displacement field $\boldsymbol{u}_{n+1}\in{}^h\mathscr{K}_{u,n+1}$ such that the body $\mathscr{B}$ is in equilibrium as stated by the Virtual Work Principle

$$\mathbf{M}\ddot{\mathbf{u}}_{n+1}+\mathbf{f}^{\text{int}}(\mathbf{u}_{n+1})-\mathbf{f}^{\text{ext}}_{n+1}=\mathbf{0},\tag{2.48}$$

where $\mathbf{f}^{\text{int}}$ e $\mathbf{f}^{\text{ext}}_{n+1}$ are the global vectors of internal and external forces defined as

$$\mathbf{f}^{\text{int}}\equiv\int_{{}^h\Omega_0}\mathbf{G}^{g\,T}\hat{\boldsymbol{P}}(\boldsymbol{F}_{n+1},\boldsymbol{\alpha}_n)\,\mathrm{d}v,\tag{2.49}$$

$$\mathbf{f}^{\text{ext}}_{n+1}\equiv\int_{{}^h\Omega_0}\mathbf{N}^{g\,T}\mathbf{b}_{0,n+1}\,\mathrm{d}v+\int_{\partial^h\Omega_{\text{traction},0}}\mathbf{N}^{g\,T}\mathbf{t}_{0,n+1}\,\mathrm{d}a.\tag{2.50}$$

and $\mathbf{M}$ is the mass matrix defined as

$$\mathbf{M}=\int_{{}^h\Omega_0}\rho_0\mathbf{N}^{g\,T}\mathbf{N}^g\,\mathrm{d}v.\tag{2.51}$$

The global vectors for the internal and external forces are usually obtained by assemblage of their elemental counterparts as

$$\mathbf{f}^{\text{int}}=\mathop{\mathsf{A}}_{e=1}^{n_{\text{elem}}}\left(\mathbf{f}^{\text{int}}\right)^{(e)},\tag{2.52}$$

$$\mathbf{f}^{\text{ext}}=\mathop{\mathsf{A}}_{e=1}^{n_{\text{elem}}}\left(\mathbf{f}^{\text{ext}}\right)^{(e)},\tag{2.53}$$

where the elemental vectors in the spatial description are defined as

$$\left(\mathbf{f}^{\text{int}}\right)^{(e)} \equiv \int_{{}^h\Omega^{(e)}} \mathbf{B}^T \hat{\boldsymbol{\sigma}}(\boldsymbol{F}_{n+1}, \boldsymbol{\alpha}_n)\, \mathrm{d}v, \tag{2.54}$$

$$\left(\mathbf{f}^{\text{ext}}_{n+1}\right)^{(e)} \equiv \int_{{}^h\Omega^{(e)}} \mathbf{N}^T \mathbf{b}_{n+1}\, \mathrm{d}v + \int_{\partial {}^h\Omega^{(e)}_{\text{traction}}} \mathbf{N}^T \mathbf{t}_{n+1}\, \mathrm{d}a, \tag{2.55}$$

and in material description as

$$\left(\mathbf{f}^{\text{int}}\right)^{(e)} \equiv \int_{{}^h\Omega^{(e)}_0} \mathbf{G}^T \hat{\boldsymbol{P}}(\boldsymbol{F}_{n+1}, \boldsymbol{\alpha}_n)\, \mathrm{d}v, \tag{2.56}$$

$$\left(\mathbf{f}^{\text{ext}}_{n+1}\right)^{(e)} \equiv \int_{{}^h\Omega^{(e)}_0} \mathbf{N}^T \mathbf{b}_{0,n+1}\, \mathrm{d}v + \int_{\partial {}^h\Omega^{(e)}_{0,\text{ traction}}} \mathbf{N}^T \mathbf{t}_{0,n+1}\, \mathrm{d}a, \tag{2.57}$$

The matrices **N**, **B**, and **G** are the elemental interpolation matrix, the symmetric elemental gradient operator, and the discrete elemental gradient operator.

In a similar manner, the global mass matrix is also usually obtained by assemblage of their elemental counterparts as

$$\mathbf{M} \equiv \mathop{\mathsf{A}}_{e=1}^{n_{\text{elem}}} \mathbf{M}^{(e)}, \tag{2.58}$$

where the elemental mass matrices in the spatial description are defined as

$$\mathbf{M}^{(e)} = \int_{{}^h\Omega^{(e)}} \rho \mathbf{N}^T \mathbf{N}\, \mathrm{d}v, \tag{2.59}$$

and the material descriptrion

$$\mathbf{M}^{(e)} = \int_{{}^h\Omega^{(e)}_0} \rho_0 \mathbf{N}^T \mathbf{N}\, \mathrm{d}v. \tag{2.60}$$

2.2.5 Numerical integration

In the Finite Element Method, the integrations over the element domain are generally performed numerically using the Gaussian Quadrature Method. Stating it's application succinctly, let $a(\boldsymbol{x})$ be a generic field, if there is a coordinate transformation from a local (or natural) normalized domain Υ to the element domain $\Omega^{(e)}$, $\boldsymbol{x}\colon \Upsilon \to \Omega^{(e)}$, the integral of $a(\boldsymbol{x})$ over the domain $\Omega^{(e)}$ can be numerically determined as

$$\int_{\Omega^{(e)}} a(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} = \int_{\Upsilon} a(\boldsymbol{x}(\boldsymbol{\zeta}))\, j(\boldsymbol{\zeta})\, \mathrm{d}\boldsymbol{\zeta} \approx \sum_{i=1}^{n_{\text{GP}}} w_i\, a(\boldsymbol{x}(\boldsymbol{\zeta}_i))\, j(\boldsymbol{\zeta}_i), \tag{2.61}$$

where $\boldsymbol{\zeta}_i$ and w_i, $i = 1, \ldots, n_{\text{GP}}$ are the positions and weigths of the Gauss sampling points in the domain Υ and $j(\boldsymbol{\zeta})$ is the determinant of the coordinate trasnformation's Jacobian defined as

$$j(\boldsymbol{\zeta}) = \det\left(\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{\zeta}}\right). \tag{2.62}$$

Note on time integration.

Chapter 3

Thermal field

For the development of the thermomechanical models, the temperature field needs to be considered. This section provides an overview of the governing equations required to describe a temperature field with the finite element method (FEM). The procedure to establish a fully discrete system of equations for the themal field is comparable to the one for the structural field in Chapter 2. Furthermore, the basics of nonlinear continuum thermodynamics have already been featured in Chapter 1. Consequently, the detailed derivation are skipped in this chapter.

3.1 Governing equations

Based on the general model presented in Section 1.5, the balance equations for the temperature field are obtained as a special case by neglecting all mechanical terms. Hence, the energy balance equation (Equation (1.42)), now in material description, reduces to

$$\rho_0 \; e = -\operatorname{div}_0 \boldsymbol{q}_0 + \rho_0 r \quad \text{in } \Omega_0, \tag{3.1}$$

where all mechanical terms are neglected. The target application of the present work are coupled generally nonlinear thermomechanical interaction problems, where the initial and the current domains are not equal, i.e. $\Omega_0 \neq \Omega$. Thus, for the sake of simplicity and in view of the later coupled problem, all following relations are expressed in material quantities. A purely thermal analysis is independent of the deformation, so that reference and current configuration are identical and the domain remains constant, i.e. $\Omega_0 \equiv \Omega$.

3.2 Thermal constitutive initial value problem

From Section 1.6, discarding all variables related to the mechanical problem, the general thermal constitutive initial value problem is

Problem 3.1 | General thermal constitutive intial value problem.

Given the initial value of the internal variables $\boldsymbol{\alpha}(t_0)$ and the history of the

temperature distribution and its gradient

$$\theta(t), \quad t \in [t_0, t_{\text{end}}], \tag{3.2}$$
$$\boldsymbol{g}_0(t), \quad t \in [t_0, t_{\text{end}}], \tag{3.3}$$

find the function for $\boldsymbol{q}_0(t)$, and $\boldsymbol{\alpha}(t)$ such that the constitutive equations

$$s = -\frac{\partial \psi}{\partial \theta}, \tag{3.4}$$
$$\psi = \psi(\theta, \boldsymbol{\alpha}), \tag{3.5}$$
$$\dot{\boldsymbol{\alpha}} = f(\theta, \boldsymbol{g}_0, \boldsymbol{\alpha}), \tag{3.6}$$
$$\frac{1}{\theta}\boldsymbol{q}_0 = g(\theta, \boldsymbol{g}_0, \boldsymbol{\alpha}). \tag{3.7}$$

are satisfied for every $t \in [t_0, t_{\text{end}}]$.

No distinction between spatial and material configurations applies as $\Omega = \Omega_0$. Next, a standard set of assumptions are introduced.

Helmholtz free energy As a first step, the specific heat C_V is established and defined according to the thermodynamical principles to be the amount of heat required to change a unit mass of a substance by one degree in temperature, i.e.

$$C_{\text{V}} = \frac{\partial e}{\partial \theta}. \tag{3.8}$$

The index $(\cdot)_{\text{V}}$ denotes that C_{V} is measured at constant volume [1]. Its dimensions are energy over temperature, i.e., $[\text{E}/\Theta]$, and using the International System of Units (SI), C_V is expressed in joule per kelvin. Using Equation 1.51, the specific heat at constant volume can be written as

$$C_{\text{V}} = -\frac{\partial^2 \psi}{\partial \theta^2}\theta = \frac{\partial s}{\partial \theta}\theta. \tag{3.9}$$

Furthermore, for the application to metals, a constant specific heat capacity (i.e. $C_{\text{V}} = \text{const.}$) is a valid assumption, utilised e.g. in **?**, **?**, **?**, and **?**.

Law for the heat flux Since deformation is not taken into account in purely thermal analyses, the material's heat flux and the heat flux in space coincide, or $\boldsymbol{q}_0 \equiv \boldsymbol{q}$. This relationship also holds true for both the material gradient and the spatial gradient, therefore $\nabla_0\theta \equiv \nabla\theta$. A constitutive law for the heat flow must be determined that links the heat flux $\boldsymbol{q}_0$ with its dual variable $\boldsymbol{g}_0$ and the temperature θ in order to meet the dissipation inequality due to conduction (Equation (1.56)). In light of this, the so-called Fourier's law, which is linear and isotropic and is stated as

$$\boldsymbol{q}_0 = -k\boldsymbol{g}_0. \tag{3.10}$$

The thermal conductivity k is considered to be constant and positive in this context, that is $k \geq 0$. Heat is therefore transferred in the direction of decreasing temperatures.

[1] According to **?** the specific heat at constant volume and constant pressure for solids can be suitable assumed to be equal.

In addition to Fourier's law, there are several constitutive laws for the heat flux that may be found in the literature. One such rule is Duhamel's law of heat conduction (see e.g. **?**), which swaps the scalar constant conductivity for a positive semi-definite second-order tensor. The conductivity tensor simplifies to $\boldsymbol{k} = k\boldsymbol{I}$ if Duhamel's law is constrained to thermally isotropic behavior (i.e. no preferred direction). Fourier's law is recovered as a particular case of Duhamel's law if a constant heat conductivity is assumed. But in the current work, Fourier's law is the only one taken into account since it is a simple law that yields physical effects.

"Standard" thermal constitutive description No extra internal variables $\boldsymbol{\alpha}$ are considered in the present description of the thermal problem. Thus, the thermal constitutive initial value problem given the standard assumptions laid out above accepts a closed form solution.

Problem 3.2 | "Standard" thermal constitutive description

Given the history of the temperature distribution and its gradient

$$\theta(t), \quad t \in [t_0, t_{\text{end}}], \tag{3.11}$$

$$\boldsymbol{g}_0(t), \quad t \in [t_0, t_{\text{end}}], \tag{3.12}$$

compute the $\boldsymbol{q}_0(t)$ and $s(t)$ at every $t \in [t_0, t_{\text{end}}]$ using the constitutive equations

$$s = -\frac{\partial \psi}{\partial \theta}, \tag{3.13}$$

$$\boldsymbol{q}_0 = -k\boldsymbol{g}_0. \tag{3.14}$$

3.3 Weak energy balance equation

The weak form of the energy balance equation must be used in order to solve the thermal issue using the FEM. The energy balance equation may be found by taking the governing equation (Equation (3.1)) in its strong form, using a variational strategy, multiplying it by the virtual temperatures ξ, and then integrating by parts.

Problem 3.3 | Weak energy balance equation

There is energy balance in the body if and only if the temperature distribution satisfies

$$\int_{\Omega_0} \left[\left(\dot{e} - \rho_0 r \right) \xi - \boldsymbol{q}_0 \cdot \nabla_0 \xi \right] \mathrm{d}v - \int_{\partial\Omega_0} h_0 \xi \, \mathrm{d}a = 0, \quad \forall \xi \in \mathscr{V}_{\theta,0}, \tag{3.15}$$

where $\mathscr{V}_{\theta,0}$ is the space of virtual temperature distributions on the body, defined by the space of sufficiently regular arbitrary temperature distributions.

3.4 The thermal initial boundary value problem

Following the same approach as in Section 2.0.3, it is now possible to introduce the the thermal initial boundary value problem. Assume that the internal variables governing

the body $\mathscr{B}$ are known at the initial time t_0. In addition, assume that the heat generated in the interior of the body is prescribed, $r(\boldsymbol{X}, t)$, $t \in [t_0, t_{\text{end}}]$, as well as,

- **Natural (or Neumann) boundary condition.** The boundary portion $\partial\Omega_{\text{heat},0}$ of $\mathscr{B}$ is subject to a prescribed history of heat flux, $h_{\text{presc},0}(\boldsymbol{X}, t) = \boldsymbol{q}_{\text{presc},0}(\boldsymbol{X}, t) \cdot \boldsymbol{m}(\boldsymbol{X})$, $\boldsymbol{X} \in \partial\Omega_{\text{heat},0}$, $t \in [t_0, t_{\text{end}}]$.
- **Essential (or Dirichlet) boundary condition.** The boundary portion $\partial\Omega_{\text{temperature},0}$ of $\mathscr{B}$ is subject to a prescribed temperature history, $\theta_{\text{presc}}(\boldsymbol{X}, t)$, $\boldsymbol{X} \in \partial\Omega_{\text{temperature},0}$, $t \in [t_0, t_{\text{end}}]$.

The body's allowable temperature distributions, as before, are all the sufficiently regular temperature fields that meet the necessary boundary condition, i.e.,

$$\mathscr{K}_\theta = \{\theta : \Omega_0 \times \mathbb{R} \to \mathbb{R} \,|\, \theta(\boldsymbol{X}, t) = \theta_{\text{presc}}(\boldsymbol{X}, t), \quad \boldsymbol{X} \in \partial\Omega_{\text{temperature},0}, \quad t \in [t_0, t_{\text{end}}]\}. \tag{3.16}$$

Heat convection and radiation are the natural boundary conditions for the thermal problem studied in this work. The prescribed heat at the boundary in the first case is given as

$$h_{\text{presc},0}(\boldsymbol{X}, t) = h_c(T(\boldsymbol{X}, t) - T_\infty), \quad \mathbf{X} \in \partial\Omega_{\text{convection},0} \subseteq \partial\Omega_{\text{heat},0} \tag{3.17}$$

with h_c denoting the coefficient of heat transfer by covection and T_∞ is the temperature of the environment. The appropriate statement for heat transmission by radiation is

$$h_{\text{presc},\,0}(\boldsymbol{X}, t) = \epsilon\sigma T^4, \quad \mathbf{X} \in \partial\Omega_{\text{radiation},0} \subseteq \partial\Omega_{\text{heat},0} \tag{3.18}$$

where $\sigma = 5.670\,373(21) \times 10^{-8}\,\text{Wm}^{-2}\text{K}^{-4}$ is Boltzman's constant and ϵ is the emissivity factor.

Using the definition of Helmholtz's energy (Equation 1.51), the balance of energy (Equation (3.1)) can be written as

$$\rho_0\dot{\psi} + \rho_0\dot{\theta}s + \rho_0\theta\dot{s} = \rho_0 r - \operatorname{div}\boldsymbol{q}_0. \tag{3.19}$$

Applying the chain rule to $\dot{s}$ yields

$$\rho_0\dot{\psi} + \rho_0\dot{\theta}s + \rho_0\theta\frac{\partial s}{\partial\theta}\dot{\theta} = \rho_0 r - \operatorname{div}\boldsymbol{q}_0. \tag{3.20}$$

Applying Equation (3.9) regarding the specific heat at constant volume reveals

$$\rho_0(\dot{\psi} + \dot{\theta}s) + \rho_0 C_V\dot{\theta} = \rho_0 r - \operatorname{div}\boldsymbol{q}_0. \tag{3.21}$$

According to Equation (1.55), the term $\dot{\psi} + \dot{\theta}s$ is precisely the internal dissipation $\mathcal{D}_{\text{int}}$, which for a strictly thermal process is zero. Accounting also for the constitutive law chosen for the heat flux, i.e., Fourier's law (Equation (3.10)), one finally obtains

$$\rho_0 C_V\dot{\theta} = -\rho_0 r + k\operatorname{div}\boldsymbol{g}_0. \tag{3.22}$$

The the weak form of the thermal constitutive initial boundary value problem can thus be stated as follows

Problem 3.4 | Thermal initial BVP.

Find an admissible temperature distribution, $\theta \in \mathscr{K}_\theta$, such that for every $t \in [t_0, t_{\text{end}}]$, the body $\mathscr{B}$ satisfies energy conservation

$$\int_{\Omega_0}\left[\left(\rho_0 C_V\dot{\theta} - \rho_0 r\right)\xi + k\boldsymbol{g}_0 \cdot \nabla_0\xi\right]\mathrm{d}v - \int_{\partial\Omega_0} h_0\xi\,\mathrm{d}a = 0, \quad \forall\xi \in \mathscr{V}_{\theta,0}, \tag{3.23}$$

where the space of virtual temperature distributions at time t is defined by

$$\mathscr{V}_{\theta,0} \equiv \left\{ \xi : \Omega_0 \to \mathbb{R} \mid \xi = 0 \quad \text{in} \quad \partial\Omega_{\text{temperature},0} \right\} \tag{3.24}$$

and at each point of $\mathscr{B}$.

3.5 Finite Element Method

Following a procedure entirely similar to the one described in Section 4.2, the global shape functions can be conveniently assembled in the so-called global interpolation vector as

$$\mathbf{N}^g(\boldsymbol{X}) \equiv \left[N_1^g(\boldsymbol{X}), N_2^g(\boldsymbol{X}), \ldots, N_{n_{\text{points}}}^g(\boldsymbol{X}) \right]. \tag{3.25}$$

The vector containg the nodal values of the temperature is denoted by θ and defined as

$$\theta(t) = \left[\theta^1(t), \ldots, \theta^{n_{\text{points}}}(t) \right]^T, \tag{3.26}$$

such that the value of the temperature inside the descretized domain ${}^h\Omega_0$ can be found from

$${}^h\theta(\boldsymbol{X}, t) \equiv \mathbf{N}^g(\boldsymbol{X})\theta(t), \quad {}^h\theta \in {}^h\mathscr{K}_\theta. \tag{3.27}$$

It is also convenient to defined the discrete gloabal gradient operator $\mathbf{H}^g$. For instance, in a 2D problem, where cartesian coordinates are employed, this discrete operator is defined as

$$\mathbf{H}^g \equiv \begin{bmatrix} \dfrac{\partial N_1^g}{\partial X} & \dfrac{\partial N_2^g}{\partial X} & \cdots & \dfrac{\partial N_{n_{\text{points}}}^g}{\partial X} \\ \dfrac{\partial N_1^g}{\partial Y} & \dfrac{\partial N_2^g}{\partial Y} & \cdots & \dfrac{\partial N_{n_{\text{points}}}^g}{\partial Y} \end{bmatrix}. \tag{3.28}$$

Applying the aforementioned finite element discretization to the thermal initial BVP yields

$$\int_{{}^h\Omega_0} \left[\left(C_V\dot{\theta} - \rho_0 r \right) \mathbf{N}^g \xi + k\boldsymbol{g}_0 \cdot \mathbf{H}^g \xi \right] \mathrm{d}v - \int_{{}^h\partial\Omega_0} h_0 \mathbf{N}^g \xi \, \mathrm{d}a = 0, \quad \forall \xi \in {}^h\mathscr{V}_{\theta,0}, \tag{3.29}$$

which can be rewritten

$$\left\{ \int_{{}^h\Omega_0} \left[(\mathbf{N}^g)^T \left(C_V\dot{\theta} - \rho_0 r \right) + k(\mathbf{H}^g)^T \mathbf{H}^g \theta \right] \mathrm{d}v - \int_{{}^h\partial\Omega_0} (\mathbf{N}^g)^T h_0 \, \mathrm{d}a \right\}^T \xi = 0, \quad \forall \xi \in {}^h\mathscr{V}_{\theta,0}, \tag{3.30}$$

where the relation $\boldsymbol{g}_0 = \mathbf{H}^g\theta$ is employed. Since Equation (3.30) must be satisfied for any $\xi \in {}^h\mathscr{V}_{\theta,0}$, the discretized thermal initial boundary value problem can be statted as

Problem 3.5 | Discretized thermal initial BVP.

Given the prescribed heat sources and heat fluxes $r(\boldsymbol{X},t)$ and $h_0(\boldsymbol{X},t)$ find the admissible nodal temperatures $\theta(t) \in {}^h\mathscr{K}_\theta$ such that the body $\mathscr{B}$ is in energetic equilibrium

$$\mathbf{C}\dot{\Theta}(t) + \mathbf{K}\Theta(t) - \mathbf{f}^{\text{ext}}(t) = \mathbf{0}, \tag{3.31}$$

where $\mathbf{C}$ and $\mathbf{K}$ are the temperature damping and stiffness matrix defined as

$$\mathbf{C} = \int_{{}^h\Omega_0} C_V \mathbf{N}^{g\,T}\mathbf{N}^g \,\mathrm{d}v., \tag{3.32}$$

$$\mathbf{K} = \int_{{}^h\Omega_0} k\mathbf{H}^{g\,T}\mathbf{H}^g \,\mathrm{d}v. \tag{3.33}$$

and $\mathbf{f}^{\text{ext}}(t)$ is the global vector of external forces defined as

$$\mathbf{f}^{\text{ext}}(t) \equiv \int_{{}^h\Omega_0} \rho\mathbf{N}^{g\,T} r(\boldsymbol{X},t)\,\mathrm{d}v + \int_{\partial{}^h\Omega_{\text{heat},0}} \mathbf{N}^{g\,T} h_0(\boldsymbol{X},t)\,\mathrm{d}a. \tag{3.34}$$

Page intentionally left blank.

Chapter 4

Thermo-mechanical problem

The full thermo-mechanical analysis is conducted in the subsequent chapter using the overall framework discussed in Chapter 1. The independent mechanical and thermal problems discussed in the Chapters 2 and 3 serve as a roadmap for the developments that come next. Since the two fields are related at every point of the domain, this coupling is referred to as a volume coupling. In contrast, surface-coupled issues, such as those involving fluid-structure interactions, encompass situations where coupling only occurs at the boundary between the fluid and the structural domain. The thermomechanical potential is the typical methodology for finite deformation thermomechanical analysis (**?**).

4.0.1 Thermo-mechanical constitutive initial value problem

The following set of equations in the material configuration set up a constitutive model based on internal variables in the complete thermo-mechanical scenario, taking into account all quantities relevant to both the mechanical and the thermal domain

$$\boldsymbol{P} = \rho_0 \frac{\partial \psi}{\partial \boldsymbol{F}}, \tag{4.1}$$

$$s = -\frac{\partial \psi}{\partial \theta}, \tag{4.2}$$

$$\psi = \psi(\boldsymbol{F}, \theta, \boldsymbol{\alpha}), \tag{4.3}$$

$$\dot{\boldsymbol{\alpha}} = f(\boldsymbol{F}, \theta, \boldsymbol{g}_0, \boldsymbol{\alpha}), \tag{4.4}$$

$$\boldsymbol{q}_0 = g(\boldsymbol{F}, \theta, \boldsymbol{g}_0, \boldsymbol{\alpha}). \tag{4.5}$$

In addition to the plastic intermediate configuration that must be taken into account after the establishment of finite strain plasticity in the general non-isothermal situation, a thermal intermediate configuration must also be considered, i.e.,

$$\boldsymbol{F} = \boldsymbol{F}^t \, \boldsymbol{F}^e \, \boldsymbol{F}^p. \tag{4.6}$$

The volume is constant during plastic deformations for the majority of metals, unlike thermal deformations, which are only volumetric. Hence, plastic deformations are assumed isochoric, $J^p = \det \boldsymbol{F}^p = 1$, yielding the following split for the jacobian of the deformation

$$J = J^t J^e. \tag{4.7}$$

Following **?**, the thermal expansion is assumed to be

$$J^t = \frac{\mathrm{d}v}{\mathrm{d}v_0} = \exp(3\alpha_T \Delta\theta). \tag{4.8}$$

Combining the last two equations, the elastic volumetric deformation J^e, can be expressed as

$$J^e = J^e(\theta) = J/J^t. \tag{4.9}$$

As a result, the additional thermal intermediate configuration can be dropped, leaving only the volumetric deformations to be considered. Elastic strains balance the body, which implicitly match the thermal strains, according to Equation (4.9), if thermal stresses develop as a result of a temperature change.

The Helmholtz free energy ψ in Equation 4.3 is expressed with respect to the reference volume, so that ψ is reformulated using potential functions,to emphasize this additive decomposition, as follows

$$\rho_0 \psi\left(\boldsymbol{F}, \theta, \boldsymbol{\alpha}\right) := \hat{\mathbb{U}}\left(J^e, \theta\right) + \hat{\mathbb{W}}(\boldsymbol{F}_{\mathrm{iso}}, \theta) + \hat{\mathbb{M}}\left(J^e, \theta\right) + \hat{\mathbb{T}}(\theta) + \hat{\mathbb{K}}\left(\boldsymbol{\alpha}, \theta\right), \tag{4.10}$$

where in contrast to the deformation gradient $\boldsymbol{F}$, the Jacobi-determinant J^e relative only to the elastic component and the isochoric deformation gradient $\boldsymbol{F}_{\mathrm{iso}}$ are applied. $\hat{\mathbb{U}}$ and $\hat{\mathbb{W}}$ can be identified with the standard hyperelastic materials potentials, with the caveat that now the material parameters can depend on the temperature, whereas $\hat{\mathbb{M}}$ describes the thermomechanical coupling potential. The potential $\hat{\mathbb{T}}$ represents the purely thermal potential. Finally, $\hat{\mathbb{K}}$ is the convex plastic potential.

Based on the potential functions, the coupling of the two fields, mechanical and thermal, can be explained:

- the temperature enters the structural field via additional thermal stresses and possibly via temperature-dependent material parameters. Herein, $\hat{\mathbb{M}}$ characterizes the thermomechanical coupling potential, leading to thermal stresses and thermal expansion and dilatation, whereas $\hat{\mathbb{K}}$ being temperature-dependent and therefore enables exemplarily von Mises plasticity combined with temperature-dependent isotropic hardening and thermal softening.

- the structure enters the thermal field via coupling terms, arising from $\hat{\mathbb{M}}$ and $\hat{\mathbb{K}}$, and thus, coupling terms as the internal dissipation $\mathcal{D}_{\mathrm{int}}$ may emerge in the thermal balance equation. Furthermore, for finite deformation thermomechanical problems, where the initial domain Ω_0 deforms to Ω, so that $\Omega \neq \Omega_0$, and a Lagrangian formulation is used, the deformation enters the thermal field additionally due to the mapping of all quantities in the balance equations to the reference configuration.

As in Chapter 3, the heat conduction law is chosen to be the Fourier heat conduction law, repeated here for the sake of clarity in the spatial description

$$\boldsymbol{q} = -k\boldsymbol{g}. \tag{4.11}$$

Applying the Piola transformation yields for the material description of the heat flux vector,

$$\boldsymbol{q}_0 = -k_0 \boldsymbol{C}^{-1} \boldsymbol{g}_0. \tag{4.12}$$

Thus, the material thermo-mechanical constitutive initial value problem can be stated as

Problem 4.1 | Material thermomechanical constitutive initial value problem.

Given the initial values of the internal variables, $\boldsymbol{\alpha}(t_0)$, the history of the deformation gradient

$$\boldsymbol{F}(t), \quad t \in [t_0, t_{\text{end}}], \tag{4.13}$$

and the history of the temperature distribution and its gradient

$$\theta(t), \quad t \in [t_0, t_{\text{end}}], \tag{4.14}$$

$$\boldsymbol{g}_0(t), \quad t \in [t_0, t_{\text{end}}], \tag{4.15}$$

find the functions for $\boldsymbol{P}(t)$, $\boldsymbol{q}_0(t)$ and $\boldsymbol{\alpha}(t)$ such that the constitutive equations

$$\boldsymbol{P} = \rho_0 \frac{\partial \psi}{\partial \boldsymbol{F}}, \tag{4.16}$$

$$s = -\frac{\partial \psi}{\partial \theta}, \tag{4.17}$$

$$\psi = \frac{1}{\rho_0}\left(\hat{\mathbb{U}}\left(J^e\right) + \hat{\mathbb{W}}(\boldsymbol{F}_{\text{iso}}) + \hat{\mathbb{M}}\left(J^e, \theta\right) + \hat{\mathbb{T}}(\theta) + \hat{\mathbb{K}}\left(\boldsymbol{\alpha}, \theta\right)\right), \tag{4.18}$$

$$\boldsymbol{q}_0 = -k_0 \boldsymbol{C}^{-1} \boldsymbol{g}_0, \tag{4.19}$$

$$\dot{\boldsymbol{\alpha}} = f(\boldsymbol{F}, \theta, \boldsymbol{g}_0, \boldsymbol{\alpha}), \tag{4.20}$$

are satisfied for every $t \in [t_0, t_{\text{end}}]$.

4.0.2 Weak equilibrium. The principle of virtual work

The strong equations that enforce the equilibrium of a body can be writen using the material description as

$$\rho_0 \ddot{\boldsymbol{u}} = \text{div}_0 \boldsymbol{P} + \boldsymbol{b}_0, \quad \text{in } \Omega_0, \tag{4.21}$$

$$\rho_0 \dot{e} = \boldsymbol{P} : \dot{\boldsymbol{F}} + \rho_0 r - \text{div}_0 \boldsymbol{q}_0, \quad \text{in } \Omega_0. \tag{4.22}$$

Following an approach entirely similar to the ones presented in Chapters 2 and 3, the weak form of the linear momentum and energy balance equations can be found to be

Problem 4.2 | Weak form of the linear momentum and energy balance equations (material version).

In a material description, the body is in mechanical equilibrium and energetic balance if and only if the First Piola-Kirchhoff stress field, $\boldsymbol{P}(t)$, the heat flux vector

$\boldsymbol{q}_0(t)$, satisfy

$$\int_{\Omega_0} [\boldsymbol{P}(t) : \nabla_0 \boldsymbol{\eta} - (\boldsymbol{b}_0(t) - \rho_0 \ddot{\boldsymbol{u}}(t)) \cdot \boldsymbol{\eta}] \, \mathrm{d}v - \int_{\partial\Omega_0} \boldsymbol{t}_0(t) \cdot \boldsymbol{\eta} \, \mathrm{d}a = 0, \quad \forall \boldsymbol{\eta} \in \mathscr{V}_{u,0}, \tag{4.23}$$

$$\int_{\Omega_0} \left[\left(\rho_0 \dot{e}(t) - \boldsymbol{P}(t) : \dot{\boldsymbol{F}}(t) - \rho_0 r(t) \right) \xi - \boldsymbol{q}_0(t) \cdot \nabla_0 \xi \right] \mathrm{d}v - \int_{\partial\Omega_0} h_0(t) \xi \, \mathrm{d}a = 0, \quad \forall \xi \in \mathscr{V}_{\theta,0}, \tag{4.24}$$

where $\mathscr{V}_{u,0}$ is the space of virtual displacement of the body, defined by the space of sufficiently regular arbitrary displacements

$$\boldsymbol{\eta} : \Omega_0 \to \mathscr{U}. \tag{4.25}$$

and $\mathscr{V}_{\theta,0}$ is the space of virtual temperature distributions of the body, defined by the space of sufficiently regular arbitrary temperature distributions

$$\xi : \Omega_0 \to \mathscr{R}. \tag{4.26}$$

4.0.3 Mechanical constitutive initial boundary value problem

It is now possible to pose the thermo-mechanical constitutive initial value problem in its weak form. Assume that a body $\mathscr{B}$ is made from a generic material, characterized by a given constitutive model, whose internal variables are known at the initial time. In addition, it is assumed that the interior of the body was subjected to a prescribed history of body forces, $\boldsymbol{b}(\boldsymbol{X}, t)$, and a prescribed history of heat sources, $r(\boldsymbol{X}, t)$, $t \in [t_0, t_{\text{end}}]$, and to the following boundary conditions:

- **Natural (or Neumann) boundary condition:**
  - **Mechanical field:** The boundary portion $\Omega_{\text{traction, 0}}$ of $\mathscr{B}$ is subjected to a prescribed history of traction forces, $\boldsymbol{t}_{\text{presc},0}(\boldsymbol{X}, t)$, $\boldsymbol{X} \in \partial\Omega_{\text{traction},0}$, $t \in [t_0, t_{\text{end}}]$,
  - **Thermal field:** The boundary portion $\partial\Omega_{\text{heat},0}$ of $\mathscr{B}$ is subject to a prescribed history of heat flux, $h_{\text{presc},0}(\boldsymbol{X}, t) = \boldsymbol{q}_{\text{presc},0}(\boldsymbol{X}, t) \cdot \boldsymbol{m}(\boldsymbol{X})$, $\boldsymbol{X} \in \partial\Omega_{\text{heat},0}$, $t \in [t_0, t_{\text{end}}]$.
- **Essential (or Dirichlet) boundary condition:**
  - **Mechanical field:** The boundary portion $\Omega_{\text{motion, 0}}$ of $\mathscr{B}$ is subjected to a prescribed displacement field history, $\boldsymbol{u}_{\text{presc}}(\boldsymbol{X}, t)$, such that

    $$\boldsymbol{\varphi}(\boldsymbol{X}, t) = \boldsymbol{X} + \boldsymbol{u}_{\text{presc}}(\boldsymbol{X}, t), \quad \boldsymbol{X} \in \partial\Omega_{\text{motion, 0}}, \quad t \in [t_0, t_{\text{end}}].$$

  - **Thermal field:** The boundary portion $\partial\Omega_{\text{temperature},0}$ of $\mathscr{B}$ is subject to a prescribed temperature history, $\theta_{\text{presc}}(\boldsymbol{X}, t)$, $\boldsymbol{X} \in \partial\Omega_{\text{temperature},0}$, $t \in [t_0, t_{\text{end}}]$.

It is also convenient to define the set of kinematically admissible displacements and admissible temperature distributions of $\mathscr{B}$ as the set of all sufficiently regular displacement and temperature functions that satisfy the corresponding essential boundary conditions (**?**),

$$\mathscr{K}_u \equiv \{\boldsymbol{u} : \Omega_0 \times \mathscr{R} \to \mathscr{U} \mid \boldsymbol{u}(\boldsymbol{X}, t) = \boldsymbol{u}_{\text{presc}}(\boldsymbol{X}, t), \quad \boldsymbol{X} \in \partial\Omega_{\text{motion},0}, \quad t \in [t_0, t_{\text{end}}]\}. \tag{4.27}$$

$$\mathscr{K}_\theta \equiv \{\theta : \Omega_0 \times \mathscr{R} \to \mathscr{R} \mid \theta(\boldsymbol{X}, t) = \theta_{\text{presc}}(\boldsymbol{X}, t), \quad \boldsymbol{X} \in \partial\Omega_{\text{temperature},0}, \quad t \in [t_0, t_{\text{end}}]\}. \tag{4.28}$$

Using the definition of Helmholtz's energy (Equation 1.51), the balance of energy (Equation (3.1)) can be written as

$$\rho_0 \dot{\psi} + \rho_0 \dot{\theta} s + \rho_0 \theta \dot{s} = \boldsymbol{P} : \dot{\boldsymbol{F}} + \rho_0 r - \operatorname{div} \boldsymbol{q}_0. \tag{4.29}$$

Applying the chain rule to $\dot{s}$ yields

$$\dot{s} = \frac{\partial s}{\partial \theta}\dot{\theta} + \frac{\partial s}{\partial J^e} : \dot{J}^e + \frac{\partial s}{\partial \boldsymbol{F}_{\text{iso}}} : \dot{\boldsymbol{F}}_{\text{iso}} + \frac{\partial s}{\partial \boldsymbol{\alpha}} * \dot{\boldsymbol{\alpha}}, \tag{4.30}$$

Applying Equation (3.9) regarding the specific heat at constant volume and defining $\mathcal{H}^{\text{ep}}$ as

$$\mathcal{H}^{\text{ep}} = -\rho_0 \theta \left(\frac{\partial^2 \hat{\mathbb{U}}}{\partial J^e \partial \theta} : \dot{J}^e + \frac{\partial^2 \hat{\mathbb{W}}}{\partial \boldsymbol{F}_{\text{iso}} \partial \theta} : \dot{\boldsymbol{F}}_{\text{iso}} + \frac{\partial^2 \hat{\mathbb{M}}}{\partial J^e \partial \theta} : \dot{J}^e + \frac{\partial^2 \hat{\mathbb{K}}}{\partial \boldsymbol{\alpha} \partial \theta} * \dot{\boldsymbol{\alpha}} \right) \tag{4.31}$$

reveals

$$\rho_0 (\dot{\psi} + \dot{\theta} s) - \boldsymbol{P} : \dot{\boldsymbol{F}} + \rho_0 C_V \dot{\theta} - \mathcal{H}^{\text{ep}} = \rho_0 r - \operatorname{div} \boldsymbol{q}_0. \tag{4.32}$$

The quantitiy denoted by $\mathcal{H}^{\text{ep}}$ is the so-called Gough-Joule effect or thermo-elastoplastic heating (or cooling) effect. According to Equation (1.55), the term $\boldsymbol{P} : \dot{\boldsymbol{F}} - \rho_0 (\dot{\psi} + \dot{\theta} s)$ is precisely the internal dissipation $\mathcal{D}_{\text{int}}$. Accounting also for the constitutive law chosen for the heat flux, i.e., Fourier's law (Equation (3.10)), one finally obtains

$$\rho_0 C_V \dot{\theta} = \rho_0 r + k \operatorname{div} \boldsymbol{g}_0 + \mathcal{D}_{\text{int}} + \mathcal{H}^{\text{ep}}. \tag{4.33}$$

So the weak form of the thermo-mechanical constitutive initial boundary value problem can be stated in a spatial description as follows and in the material description as

Problem 4.3 | Material thermo-mechanical initial BVP.

Find a kinematically admissible displacement function, $\boldsymbol{u} \in \mathscr{K}_u$, and an admissible temperature distribution, $\theta \in \mathscr{K}_\theta$, such that for every $t \in [t_0, t_{\text{end}}]$, the body $\mathscr{B}$ is in mechanical and energetic equilibrium

$$\int_{\Omega_0} [\boldsymbol{P}(t) : \nabla_0 \boldsymbol{\eta} - (\boldsymbol{b}_0(t) - \rho_0 \ddot{\boldsymbol{u}}(t)) \cdot \boldsymbol{\eta}] \, \mathrm{d}v - \int_{\partial\Omega_0} \boldsymbol{t}_0(t) \cdot \boldsymbol{\eta} \, \mathrm{d}a = 0, \quad \forall \boldsymbol{\eta} \in \mathscr{V}_{u,0}, \tag{4.34}$$

$$\int_{\Omega_0} \left[\left(\rho_0 C_V \dot{\theta}(t) - \mathcal{D}_{\text{int}}(t) - \mathcal{H}^{\text{ep}}(t) - \rho_0 r(t) \right) \xi + k \boldsymbol{g}_0(t) \cdot \nabla_0 \xi \right] \mathrm{d}v - \int_{\partial\Omega_0} h_0(t) \xi \, \mathrm{d}a = 0, \quad \forall \xi \in \mathscr{V}_{\theta,0}, \tag{4.35}$$

where the space of virtual displacements at time t is defined by

$$\mathscr{V}_{u,0} \equiv \{\boldsymbol{\eta} : \Omega_0 \to \mathscr{U} \mid \boldsymbol{\eta} = \mathbf{0} \quad \text{in} \quad \partial\Omega_{\text{motion},0}\}, \tag{4.36}$$

the space of virtual temperatures at time t is defined by

$$\mathscr{V}_{\theta,0} \equiv \left\{\xi : \Omega_0 \to \mathscr{R} \mid \xi = 0 \quad \text{in} \quad \partial\Omega_{\text{temperature},0}\right\}, \tag{4.37}$$

and at each point of $\mathscr{B}$, the First Piola-Kirchhoff stress tensor is the solution of material thermomechanical constitutive initial value problem.

4.1 Time descretization

In the context of thermo-mechanical problems, a general path-dependent model is one that considers both the instantaneous deformation and temperature states as well as their history. Under these circumstances, for complex strain, $\boldsymbol{F}(t)$, or temperature paths, $\theta(t)$, the solution of the constitutive initial value problem for a given set of initial conditions is typically unknown. Therefore, it is necessary to employ a suitable numerical approach to integrate the rate constitutive equations, just as it would be in the case of a strictly mechanical problem.

Attending to the thermo-mechanical constitutive initial boundary value problem and considering the time increment $[t_n, t_{n+1}]$, this approach is comprised by the following requirements:

- **First Piola-Kirchhoff stress tensors.** Considering a time increment $[t_n, t_{n+1}]$ and given the set $\boldsymbol{\alpha}_n$ of internal variables at t_n, the deformation gradient $\boldsymbol{F}_{n+1}$ and the temperature distribution θ_{n+1} at time t_{n+1} determines the First Piola-Kirchhoff stress tensor $\boldsymbol{P}_{n+1}$ uniquely as

  $$\boldsymbol{P}_{n+1} = \hat{\boldsymbol{P}}(\boldsymbol{F}_{n+1}, \theta_{n+1}, \boldsymbol{\alpha}_n), \tag{4.38}$$

  where $\hat{\boldsymbol{P}}$ is the incremental constitutive function for the First Piola-Kirchhoff stress tensor.

- **Set of internal variables.** Assuming that the set of internal variables $\boldsymbol{\alpha}_n$ is known at t_n, the set of internal variables must be uniquely determined by the prescribed deformation gradient $\boldsymbol{F}_{n+1}$ and temperature distribution θ_{n+1} prescribed at t_{n+1} as

  $$\boldsymbol{\alpha}_{n+1} = \hat{\boldsymbol{\alpha}}(\boldsymbol{F}_{n+1}, \theta_{n+1}, \boldsymbol{\alpha}_n), \tag{4.39}$$

  where $\hat{\boldsymbol{\alpha}}$ is the incremental constitutive function for the set of internal variables.

- **Mechanical dissipation.** Considering a time increment $[t_n, t_{n+1}]$ and given the set $\boldsymbol{\alpha}_n$ of internal variables at t_n, the deformation gradient $\boldsymbol{F}_{n+1}$ and the temperature distribution θ_{n+1} at time t_{n+1} determines the mechanical dissipation $\mathcal{D}_{\text{int}}$ as

$$\mathcal{D}_{\text{int},n+1} = \hat{\mathcal{D}}_{\text{int}}(\boldsymbol{F}_{n+1}, \theta_{n+1}, \boldsymbol{\alpha}_n). \tag{4.40}$$

- **Gough-Joule effect.** Considering a time increment $[t_n, t_{n+1}]$ and given the set $\boldsymbol{\alpha}_n$ of internal variables at t_n, the deformation gradient $\boldsymbol{F}_{n+1}$ and the temperature distribution θ_{n+1} at time t_{n+1} determines the Gough-Joule effect $\mathcal{H}^{\text{ep}}$ as

$$\mathcal{H}^{\text{ep}}_{n+1} = \hat{\mathcal{H}}^{\text{ep}}(\boldsymbol{F}_{n+1}, \theta_{n+1}, \boldsymbol{\alpha}_n). \tag{4.41}$$

Generally, the numerical constitutive laws are nonlinear and path-independent within one increment.

Making use of the aforementioned time discretization, one can state the weak form of the mechanical constitutive initial boundary value problem in the material description as

Problem 4.4 | Material incremental thermo-mechanical initial BVP.

Given the set of internal variables $\boldsymbol{\alpha}_n$ at t_n, the prescribed body and traction force fields $\boldsymbol{b}_{0,n+1}$ and $\boldsymbol{t}_{0,n+1}$, as well as, the prescribed heat sources, $\boldsymbol{r}_{n+1}$ and heat fluxes, $h_{0,n+1}$, at t_{n+1}, and the prescribed deformation gradient $\boldsymbol{F}_{n+1}$ and temperature distribution θ_{n+1} at t_{n+1}, find the kinematically admissible displacement field $\boldsymbol{u}_{n+1} \in \mathscr{K}_{u,n+1}$ and the admissible temperature distribution $\theta_{n+1} \in \mathscr{K}_{\theta,n+1}$ such that the body $\mathscr{B}$ is in mechanical and energetic equilibrium

$$\int_{\Omega_0} [\hat{\boldsymbol{P}}(\boldsymbol{F}(\boldsymbol{u}_{n+1}), \theta_{n+1}, \boldsymbol{\alpha}_n) : \nabla_0 \boldsymbol{\eta} - (\boldsymbol{b}_{0,n+1} - \rho_0 \ddot{\boldsymbol{u}}_{n+1}) \cdot \boldsymbol{\eta}] \, \mathrm{d}v - \int_{\partial\Omega_0} \boldsymbol{t}_{0,n+1} \cdot \boldsymbol{\eta} \, \mathrm{d}a = 0, \quad \forall \boldsymbol{\eta} \in \mathscr{V}_{u,n+1}, \tag{4.42}$$

$$\int_{\Omega_0} \left[\left(\rho_0 C_{\text{V}} \dot{\theta}_{n+1} - \hat{\mathcal{D}}_{\text{int}}(\boldsymbol{F}(\boldsymbol{u}_{n+1}), \theta_{n+1}, \boldsymbol{\alpha}_n) - \hat{\mathcal{H}}^{\text{ep}}(\boldsymbol{F}(\boldsymbol{u}_{n+1}), \theta_{n+1}, \boldsymbol{\alpha}_n) - \rho_0 r_{n+1} \right) \xi - \boldsymbol{g}_{0,n+1} \cdot \nabla_0 \xi \right] \mathrm{d}v - \int_{\partial\Omega_0} \hat{h}_{0,n+1} \xi \, \mathrm{d}a = 0, \quad \forall \xi \in \mathscr{V}_{\theta,n+1}, \tag{4.43}$$

where

$$\mathscr{K}_{u,n+1} \equiv \{\boldsymbol{u} : \Omega \times \mathscr{R} \to \mathscr{U} \mid \boldsymbol{u}_{n+1}(\boldsymbol{X}) = \boldsymbol{u}_{\text{presc},n+1}(\boldsymbol{X}),\ \boldsymbol{X} \in \partial\Omega_{\text{motion},0}\}, \tag{4.44}$$

$$\mathscr{K}_{\theta,n+1} \equiv \{\boldsymbol{u} : \Omega \times \mathscr{R} \to \mathscr{R} \mid \theta_{n+1}(\boldsymbol{X}) = \theta_{\text{presc},n+1}(\boldsymbol{X}),\ \boldsymbol{X} \in \partial\Omega_{\text{temperature},0}\}. \tag{4.45}$$

4.2 Finite Element Method

It is now possible to apply the finite element method to discretize spatial the incremental thermo-mechanical initial boundary value problem. The approach is entirely similar to the one present in the context of the strictly mechanical (see Section 4.2) and strictly thermal (see Section 3.5) problems. As such, some details are omitted.

4.2.1 Interpolation

Defining the global vector of nodal displacements $\mathbf{u}$ and the global vector of nodal temperatures Θ, like before, as

$$\mathbf{u}(t) = \left[u_1^1(t), \dots, u_{n_{\text{dim}}}^1(t), \dots, u_1^{n_{\text{points}}}(t), \dots, u_{n_{\text{dim}}}^{n_{\text{points}}}(t) \right]^T, \tag{4.46}$$

$$\Theta(t) = \left[\theta^1(t), \theta^2(t), \dots, \theta^{n_{\text{points}}}(t) \right]^T, \tag{4.47}$$

the displacement $\boldsymbol{u}(\boldsymbol{X}, t)$ and temperature $\theta(\boldsymbol{X}, t)$ fields defined over the global domain Ω_0, can be found at any point $\boldsymbol{X}$ as

$$^h\boldsymbol{u}(\boldsymbol{X}, t) \equiv \mathbf{N}^{g,u}(\boldsymbol{X})\mathbf{u}(t), \quad ^h\boldsymbol{u} \in {}^h\mathscr{K}_u, \tag{4.48}$$

$$^h\theta(\boldsymbol{X}, t) \equiv \mathbf{N}^{g,\theta}(\boldsymbol{X})\Theta(t), \quad ^h\boldsymbol{\theta} \in {}^h\mathscr{K}_\theta, \tag{4.49}$$

where $\mathbf{N}^{g,u}$ and $\mathbf{N}^{g,\theta}$ are the global interpolation matrices with appropriate dimensions given the dimensions of $\mathbf{u}$ and Θ. The global shape functions used to interpolate between the nodal values of the displacement and the temperature can even be different if need be.

4.2.2 Spatial discretization

The application of finite element discretization to the mechanical part of the incremental thermo-mechanical constitutive initial boundary value problem is exactly the same as the one presented in Section 2.2.4 for the strictly mechanical problem, as the the equation to be discretized is the same On the other hand, applying the discretization to the thermal part of the incremental thermo-mechanical constitutive initial boundary value problem, yields

$$\int_{^h\Omega_0} \left[\left(\rho_0 C_V \mathbf{N}^{g,\theta} \dot{\Theta}_{n+1} - \hat{\mathcal{D}}_{\text{int}} - \hat{\mathcal{H}}^{\text{ep}} - \rho_0 r_{n+1} \right) \cdot \mathbf{N}^{g,\theta} \xi - \mathbf{H}^{g,\theta} \Theta_{n+1} \cdot \mathbf{H}^{g,\theta} \xi \right] \mathrm{d}v - \int_{^h\partial\Omega_0} h_{0,n+1} \mathbf{N}^{g,\theta} \xi \, \mathrm{d}a = 0, \quad \forall \xi \in {}^h\mathscr{V}_\theta, \tag{4.50}$$

where $\mathbf{H}^{g,\theta}$ is the discrete global gradient operator for scalars, defined in Equation (3.28).

Equation (4.51) can be rewritten as

$$\left\{ \int_{^h\Omega_0} \left[\mathbf{N}^{g,\theta\,T} \left(\rho_0 C_V \mathbf{N}^{g,\theta} \dot{\Theta}_{n+1} - \hat{\mathcal{D}}_{\text{int}} - \hat{\mathcal{H}}^{\text{ep}} - \rho_0 r_{n+1} \right) - \mathbf{H}^{g,\theta\,T} \mathbf{H}^{g,\theta} \Theta_{n+1} \right] \mathrm{d}v - \int_{\partial^h\Omega_0} \mathbf{N}^{g,\theta\,T} h_{0,n+1} \, \mathrm{d}a \right\}^T \xi = 0, \quad \forall \xi \in {}^h\mathscr{V}_\theta, \tag{4.51}$$

and, since it must be satisfied for any $\xi \in {}^h\mathscr{V}_\theta$, the incremental discretized thermo-mechanical constitutive initial boundary value problem can thus be stated in the material description as

Problem 4.5 | Material incremental discretized thermo-mechanical initial BVP.

Given the set of internal variables $\boldsymbol{\alpha}_n$ at t_n, the prescribed body and traction force fields $\boldsymbol{b}_{0,n+1}$ and $\boldsymbol{t}_{0,n+1}$, as well as, the prescribed heat sources and heat flux fields $\boldsymbol{r}_{0,n+1}$ and $h_{0,n+1}$, and both the prescribed deformation gradient $\boldsymbol{F}_{n+1}$ and the prescribed temperature θ_{n+1} at t_{n+1}, find the kinematically admissible nodal displacement field $\boldsymbol{u}_{n+1} \in {}^h\mathscr{K}_{u,n+1}$ and the admissible nodal temperature field $\theta_{n+1} \in {}^h\mathscr{K}_{\theta,n+1}$ such that the body $\mathscr{B}$ is in mechanical and energetic equilibrium

$$\mathbf{M}\ddot{\mathbf{u}}_{n+1} + \mathbf{f}_u^{\text{int}}(\Theta_{n+1}, \mathbf{u}_{n+1}) - \mathbf{f}_{u,n+1}^{\text{ext}} = \mathbf{0}, \tag{4.52}$$

$$\mathbf{C}\dot{\Theta}_{n+1} + \mathbf{K}\Theta_{n+1} + \mathbf{f}_\theta^{\text{int}}(\Theta_{n+1}, \mathbf{u}_{n+1}) - \mathbf{f}_{\theta,n+1}^{\text{ext}} = \mathbf{0}, \tag{4.53}$$

where $\mathbf{f}_u^{\text{int}}$ e $\mathbf{f}_{u,n+1}^{\text{ext}}$ are the mechanical global vectors of internal and external forces defined as

$$\mathbf{f}_u^{\text{int}} \equiv \int_{{}^h\Omega_0} \mathbf{G}^{g,u\,T} \hat{\boldsymbol{P}}(\boldsymbol{F}(\boldsymbol{u}_{n+1}), \theta_{n+1}\boldsymbol{\alpha}_n)\, \mathrm{d}v, \tag{4.54}$$

$$\mathbf{f}_{u,n+1}^{\text{ext}} \equiv \int_{{}^h\Omega_0} \mathbf{N}^{g,u\,T} \mathbf{b}_{0,n+1}\, \mathrm{d}v + \int_{\partial^h\Omega_{\text{traction},0}} \mathbf{N}^{g,u\,T} \mathbf{t}_{0,n+1}\, \mathrm{d}a, \tag{4.55}$$

and $\mathbf{f}_\theta^{\text{int}}$ e $\mathbf{f}_{\theta,n+1}^{\text{ext}}$ the thermal global vectors of internal and external forces defined as

$$\mathbf{f}_\theta^{\text{int}} \equiv \left[\int_{{}^h\Omega_0} \mathbf{N}^{g,\theta\,T} \hat{\mathcal{D}}_{\text{int}}(\boldsymbol{F}(\boldsymbol{u}_{n+1}), \theta_{n+1}\boldsymbol{\alpha}_n) + \mathbf{N}^{g,\theta\,T} \hat{\mathcal{H}}^{\text{ep}}(\boldsymbol{F}(\boldsymbol{u}_{n+1}), \theta_{n+1}\boldsymbol{\alpha}_n) \right] \mathrm{d}v, \tag{4.56}$$

$$\mathbf{f}_{\theta,n+1}^{\text{ext}} \equiv \int_{{}^h\Omega_0} \mathbf{N}^{g,\theta\,T} \mathbf{r}_{0,n+1}\, \mathrm{d}v + \int_{\partial^h\Omega_{\text{heat},0}} \mathbf{N}^{g,\theta\,T} h_{0,n+1}\, \mathrm{d}a. \tag{4.57}$$

The mass matrix $\mathbf{M}$ and the thermal capacitance matrix $\mathbf{C}$ are defined by

$$\mathbf{M} = \int_{{}^h\Omega_0} \rho_0 \mathbf{N}^{g,u\,T} \mathbf{N}^{g,u}\, \mathrm{d}v, \tag{4.58}$$

$$\mathbf{C} = \int_{{}^h\Omega_0} \rho_0 C_{\text{V}} \mathbf{N}^{g,\theta\,T} \mathbf{N}^{g,\theta}\, \mathrm{d}v, \tag{4.59}$$

$$\mathbf{K} = \int_{{}^h\Omega_0} \mathbf{H}^{g,\theta\,T} \mathbf{H}^{g,\theta}\, \mathrm{d}v. \tag{4.60}$$

The global vectors for the internal and external forces are usually obtained by assemblage of their elemental counterparts as

$$\mathbf{f}_u^{\text{int}} = \mathop{\mathsf{A}}_{e=1}^{n_{\text{elem}}} \left(\mathbf{f}_u^{\text{int}}\right)^{(e)}, \tag{4.61}$$

$$\mathbf{f}_u^{\text{ext}} = \mathop{\mathsf{A}}_{e=1}^{n_{\text{elem}}} \left(\mathbf{f}_u^{\text{ext}}\right)^{(e)}, \tag{4.62}$$

$$\mathbf{f}_\theta^{\text{int}} = \mathop{\mathsf{A}}_{e=1}^{n_{\text{elem}}} \left(\mathbf{f}_\theta^{\text{int}}\right)^{(e)}, \tag{4.63}$$

$$\mathbf{f}_\theta^{\text{ext}} = \mathop{\mathsf{A}}_{e=1}^{n_{\text{elem}}} \left(\mathbf{f}_\theta^{\text{ext}}\right)^{(e)}, \tag{4.64}$$

where the mechanical elemental vectors in the material description are defined as

$$\mathbf{f}_u^{\text{int}} \equiv \int_{{}^h\Omega_0^{(e)}} \mathbf{G}^{u\,T} \hat{\boldsymbol{P}}(\boldsymbol{F}(\boldsymbol{u}_{n+1}), \theta_{n+1}\boldsymbol{\alpha}_n)\, \mathrm{d}v, \tag{4.65}$$

$$\mathbf{f}_{u,n+1}^{\text{ext}} \equiv \int_{{}^h\Omega_0^{(e)}} \mathbf{N}^{u\,T} \mathbf{b}_{0,n+1}\, \mathrm{d}v + \int_{\partial {}^h\Omega_{\text{traction},0}^{(e)}} \mathbf{N}^{u\,T} \mathbf{t}_{0,n+1}\, \mathrm{d}a, \tag{4.66}$$

and the thermal vectors, also in the material description, are

$$\mathbf{f}_\theta^{\text{int}} \equiv \left[\int_{{}^h\Omega_0^{(e)}} \mathbf{N}^{\theta\,T} \hat{\mathcal{D}}_{\text{int}}(\boldsymbol{F}(\boldsymbol{u}_{n+1}), \theta_{n+1}\boldsymbol{\alpha}_n) + \mathbf{N}^{\theta\,T} \hat{\mathcal{H}}^{\text{ep}}(\boldsymbol{F}(\boldsymbol{u}_{n+1}), \theta_{n+1}\boldsymbol{\alpha}_n)\right] \mathrm{d}v, \tag{4.67}$$

$$\mathbf{f}_{\theta,n+1}^{\text{ext}} \equiv \int_{{}^h\Omega_0^{(e)}} \mathbf{N}^{\theta\,T} \mathbf{r}_{0,n+1}\, \mathrm{d}v + \int_{\partial {}^h\Omega_{\text{heat},0}^{(e)}} \mathbf{N}^{\theta\,T} h_{0,n+1}\, \mathrm{d}a. \tag{4.68}$$

The elemental mechanical interpolation of the matrices $\mathbf{N}^u$, $\mathbf{B}^u$, $\mathbf{N}^\theta$ and $\mathbf{H}^\theta$ are the elemental mechanical interpolation matrix, the mechanical discrete elemental gradient operator, the elemental thermal interpolation matrix, and the thermal discrete elemental gradient operator for scalars.

In a similar manner, the global mass and capacitance matrices are also usually obtained by assemblage of their elemental counterparts as

$$\mathbf{M} \equiv \mathop{\mathsf{A}}_{e=1}^{n_{\text{elem}}} \mathbf{M}^{(e)}, \tag{4.69}$$

$$\mathbf{C} \equiv \mathop{\mathsf{A}}_{e=1}^{n_{\text{elem}}} \mathbf{C}^{(e)}, \tag{4.70}$$

$$\mathbf{K} \equiv \mathop{\mathsf{A}}_{e=1}^{n_{\text{elem}}} \mathbf{K}^{(e)}, \tag{4.71}$$

where the elemental mass matrices in the material description are defined as

$$\mathbf{M}^{(e)} = \int_{h\Omega^{(e)}} \rho_0 \mathbf{N}^{u\,T} \mathbf{N}^u \,\mathrm{d}v, \tag{4.72}$$

the elemental thermal capacitance matrices as

$$\mathbf{C}^{(e)} = \int_{h\Omega^{(e)}} \rho_0 C_\mathrm{V} \mathbf{N}^{\theta\,T} \mathbf{N}^\theta \,\mathrm{d}v, \tag{4.73}$$

and the elemental thermal stiffness matrix as

$$\mathbf{K}^{(e)} = \int_{h\Omega^{(e)}} \mathbf{H}^{\theta\,T} \mathbf{H}^\theta \,\mathrm{d}v. \tag{4.74}$$

Chapter 5

Validation results for the thermal solver

This chapter provides validation results for the thermal solver implemented in this work. The appropriate examples are sourced from **?** - Prüfung und Validierung von Rechenprogramm für Brandschutznachweise mittels allgemeiner Rechenverfahren and the linear thermo-elastic test in the **?**. They include thermal effects such as variable conductivity, heat convection and radiation at the boundary. The numerical solutions are obtained using the thermal solver in LINKS, employing TRI3, TRI6, QUAD4, QUAD8 elements, in two-dimensions, and TETRA4 and TETRA10 elements in three-dimensions. No convergence study was performed, however the mesh size was chosen small enough so that assuming convergence of the FEM solution is reasonable. Moreover the good agreement with reference solutions supports this assumption.

5.1 Validation example 1 - DIN EN 1991-1-2/NA:2010-12: Anhang CC - Prüfung und Validierung von Rechenprogramm für Brandschutznachweise mittels allgemeiner Rechenverfahren - Beispiel 1)

5.1.1 Description

The geometry examined is a square plate with side length equal to 1 m, as shown in Figure 5.1. The boundary conditions considered are as follows: the left, upper and right edges are assumed to be adiabatic. At the lower edge there is heat transfer by convection with a heat convection coefficient h_c equation to $1\,\mathrm{Wm^{-2}K^{-1}}$ and an environment temperature equal to 0 °C (see Equation 3.17). The initial temperature for the entire plate is 1000 °C. Reference values for the temperature at the middle of the upper edge are supplied to determine performance. The relevant properties of the material making up the plate are its conductivity k, equal to $1\,\mathrm{Wm^{-1}K^{-1}}$, its specific heat c_p, equal to $1\,\mathrm{Jkg^{-1}K^{-1}}$, and its density ρ, set equal to $1000\,\mathrm{kgm^{-3}}$. Table 5.1 summarizes all the information regarding initial and boundary conditions, geometry and material properties.

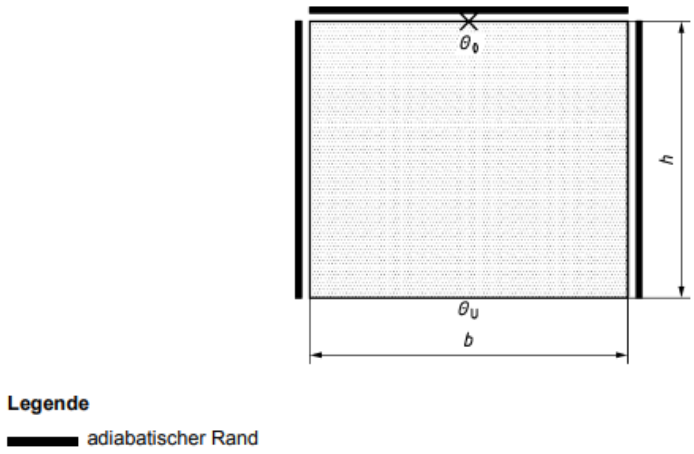
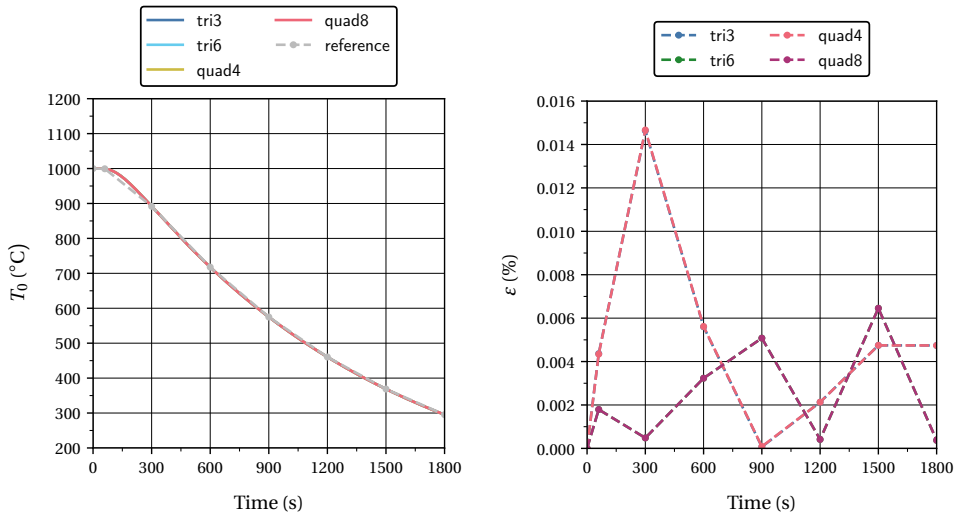


Figure 5.1: Geometry and boundary conditions considered in the validation example 1 (**?**).

Table 5.1: Material properties, and initial and boundary conditions for validation example 1.

Material Properties			Effective value
Conductivity	k	($\mathrm{Wm^{-1}K^{-1}}$)	1
Specific heat	c_p	($\mathrm{Jkg^{-1}K^{-1}}$)	1
Density	ρ	($\mathrm{kg/m^3}$)	1.000×10^3
Boundary Conditions			
Dimensions	h, b	(m)	1
Heat convection coefficient	h_c	($\mathrm{W/m^2/K}$)	1
Initial Conditions			
Ambient temperature	T_∞	(°C)	0
Temperature in cross-section	T_0	(°C)	1.000×10^3
Reference value			
Temperature T at point X	(°C)		



(a) (b)

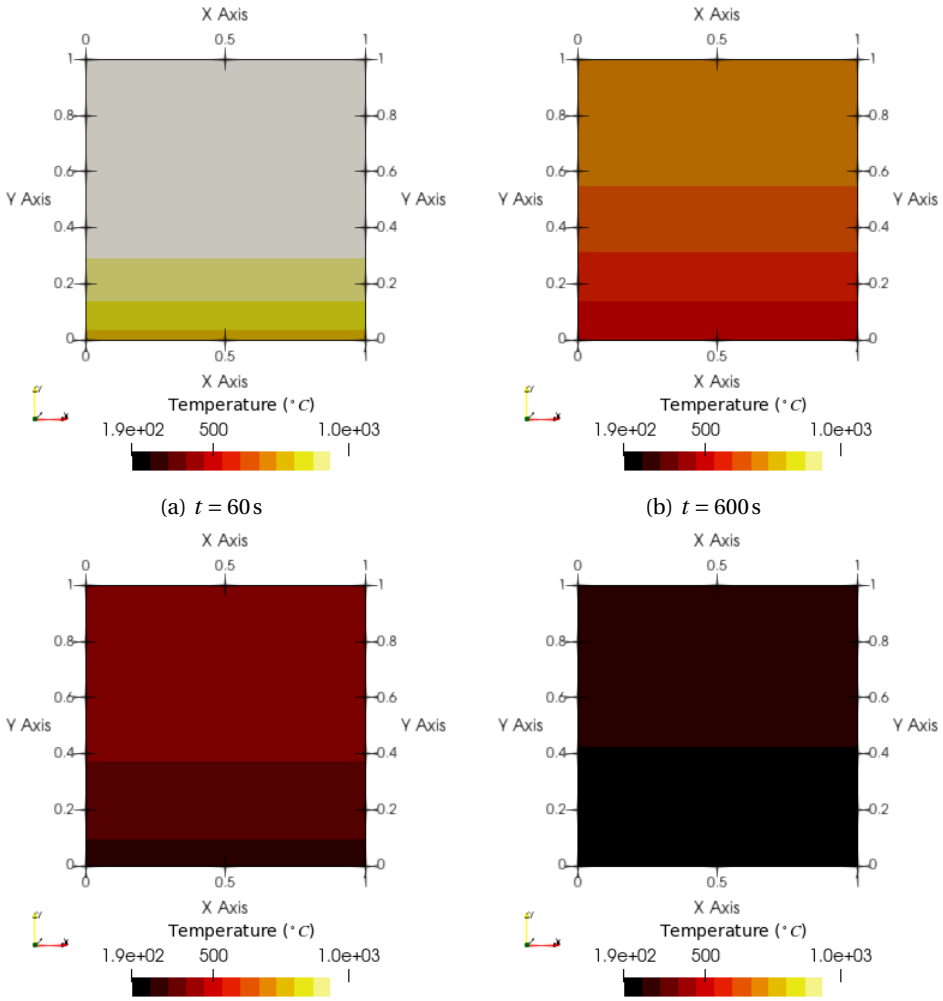
Figure 5.2: Numerical results for the validation example 1. (a) Temperature values at X as a function of time. (b) Relative error in percentage as function of time.

5.1.2 Results

The numerical solutions obtained using FEM are presented in Table 5.2, as well as, the reference values and the corresponding relative difference. Figure 5.2 presents the same results in graphical form. It can seen that that agreement between the numerical results and the reference solutions is very good, with relative error always smaller than 0.02%. recomends a relative difference ±1% and an absolute difference ±5°C. Figure 5.3 shows different time instants of the numerical solution using TRI3 elements. The evolution of the temperature field depicted seems reasonable given the description of the problem.

Table 5.2: Reference and computed values for T_0 concerning the validation example 1.

Time (s)	Reference value T_0 (°C)	Element Type	Computed value T_0' (°C)	Relative difference ε (%)
0	1000.0	TRI3	1000.0000	0.00
		TRI6	1000.0000	0.00
		QUAD4	1000.0000	0.00
		QUAD8	1000.0000	0.00
60	999.3	TRI3	999.3436	4.36×10^{-3}
		TRI6	999.2821	1.79×10^{-3}
		QUAD4	999.3434	4.34×10^{-3}
		QUAD8	999.2821	1.79×10^{-3}
300	891.8	TRI3	891.9305	1.46×10^{-2}
		TRI6	891.7957	4.82×10^{-4}
		QUAD4	891.9308	1.47×10^{-2}
		QUAD8	891.7957	4.82×10^{-4}
600	717.7	TRI3	717.7402	5.60×10^{-3}
		TRI6	717.6768	3.23×10^{-3}
		QUAD4	717.7403	5.62×10^{-3}
		QUAD8	717.6768	3.23×10^{-3}
900	574.9	TRI3	574.9004	6.96×10^{-5}
		TRI6	574.8708	5.08×10^{-3}
		QUAD4	574.9005	8.70×10^{-5}
		QUAD8	574.8708	5.08×10^{-3}
1200	460.4	TRI3	460.4098	2.13×10^{-3}
		TRI6	460.4019	4.13×10^{-4}
		QUAD4	460.4098	2.13×10^{-3}
		QUAD8	460.4019	4.13×10^{-4}
1500	368.7	TRI3	368.7175	4.75×10^{-3}
		TRI6	368.7238	6.46×10^{-3}
		QUAD4	368.7175	4.75×10^{-3}
		QUAD8	368.7238	6.46×10^{-3}
1800	295.3	TRI3	295.2860	4.74×10^{-3}
		TRI6	295.3011	3.73×10^{-4}
		QUAD4	295.2860	4.74×10^{-3}
		QUAD8	295.3011	3.73×10^{-4}



(a) $t = 60\,\mathrm{s}$ (b) $t = 600\,\mathrm{s}$

(c) $t = 1200\,\mathrm{s}$ (d) $t = 1800\,\mathrm{s}$

Figure 5.3: Numerical results regarding the evolution of the temperature distribution for the validation example 1 using a TRI3 mesh.

5.2 Validation example 2 - DIN EN 1991-1-2/NA:2010-12: Anhang CC - Prüfung und Validierung von Rechenprogramm für Brandschutznachweise mittels allgemeiner Rechenverfahren - Beispiel 2)

5.2.1 Description

The geometry examined is a square plate with side length equal to 0.2 m, as shown in Figure 5.4. There is heat transfer by convection along all the edges with a heat convection coefficient h_c equation to $10\,\mathrm{Wm^{-2}K^{-1}}$ and an environment temperature equal to 1000 °C (see Equation 3.17). Thre is also heat transfer through radition, with the emissivity $\varepsilon_{\mathrm{res}}$ equal to 0.8. The initial temperature for the entire plate is 0 °C. Reference values for the temperature in the middle of the plate are supplied to determine performance. The relevant properties of the material making up the plate are its conductivity k, which follows a linear behavior (see Table 5.3), its specific heat c_p, equal to $1000\,\mathrm{Jkg^{-1}K^{-1}}$, and its density ρ, set equal to $2400\,\mathrm{kgm^{-3}}$. Table 5.3 summarizes all the information regarding initial and boundary conditions, geometry and material properties.

5.2.2 Results

The numerical solutions obtained using FEM are presented in Table 5.4, as well as, the reference values and the corresponding relative difference. Figure 5.5 presents the same results in graphical form. recomends for $t \leq 60\,\mathrm{min}$ an absolute difference smaller than ±5 °C, and for $t > 60\,\mathrm{min}$, a relative difference smaller than ±2%. It can be seen that that agreement between the numerical results and the reference solutions is acceptable. For $t \leq 60\,\mathrm{min}$ the linear elements do not satisfy the recomendation set forth by . Otherwise the requirements are completly fullfiled. Figure 5.6 shows different time instants of the numerical solution using TRI3 elements. The evolution of the temperature field depicted seems reasonable given the description of the problem.

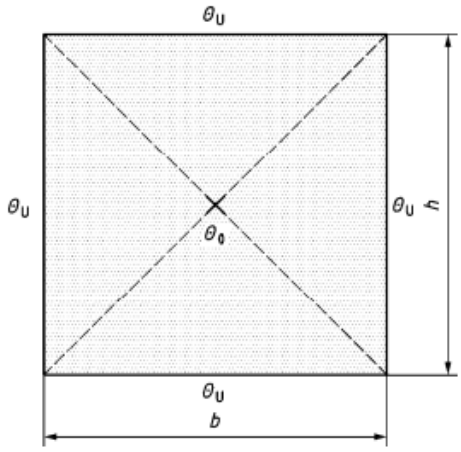


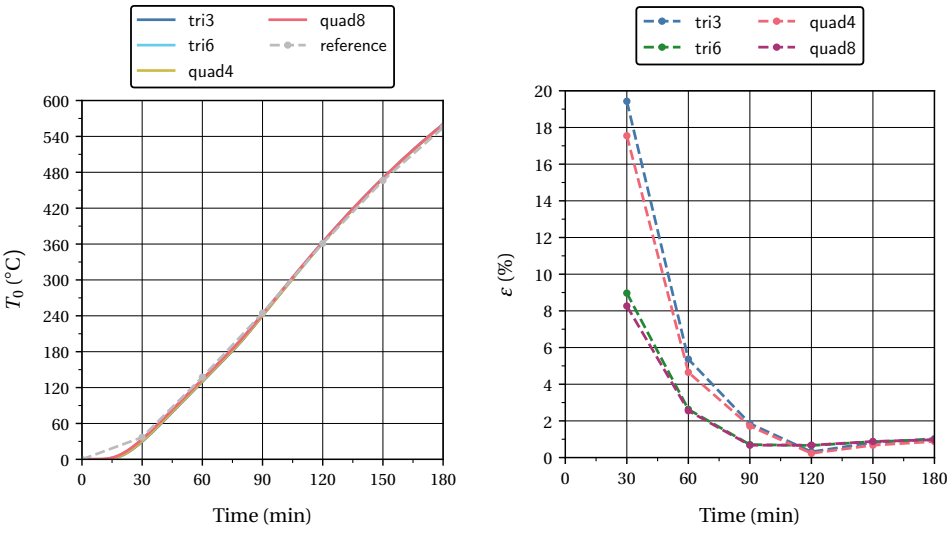
Figure 5.4: Geometry and boundary conditions considered in the validation example 2. (**?**)

Table 5.3: Material properties, and initial and boundary conditions for validation example 2.

Material Properties			Effective value	
			T	$\lambda(T)$
Conductivity (Linear behavior)	k	($\mathrm{W m^{-1} K^{-1}}$)	0	1.5
			200	0.7
			1000	0.5
Specific heat	c_p	($\mathrm{J kg^{-1} K^{-1}}$)	1000	
Density	ρ	($\mathrm{kg/m^3}$)	2400	
Boundary Conditions				
Dimensions	h, b	(m)	0.2	
Heat convection coefficient	h_c	($\mathrm{W/m^2/K}$)	10	
Emissivity	$\varepsilon_{\mathrm{res}}$		0.8	
Initial Conditions				
Ambient temperature	T_∞	(°C)	1000	
Temperature in cross-section	T_0	(°C)	0	
Reference value				
Temperature T_0 in point X		(°C)		

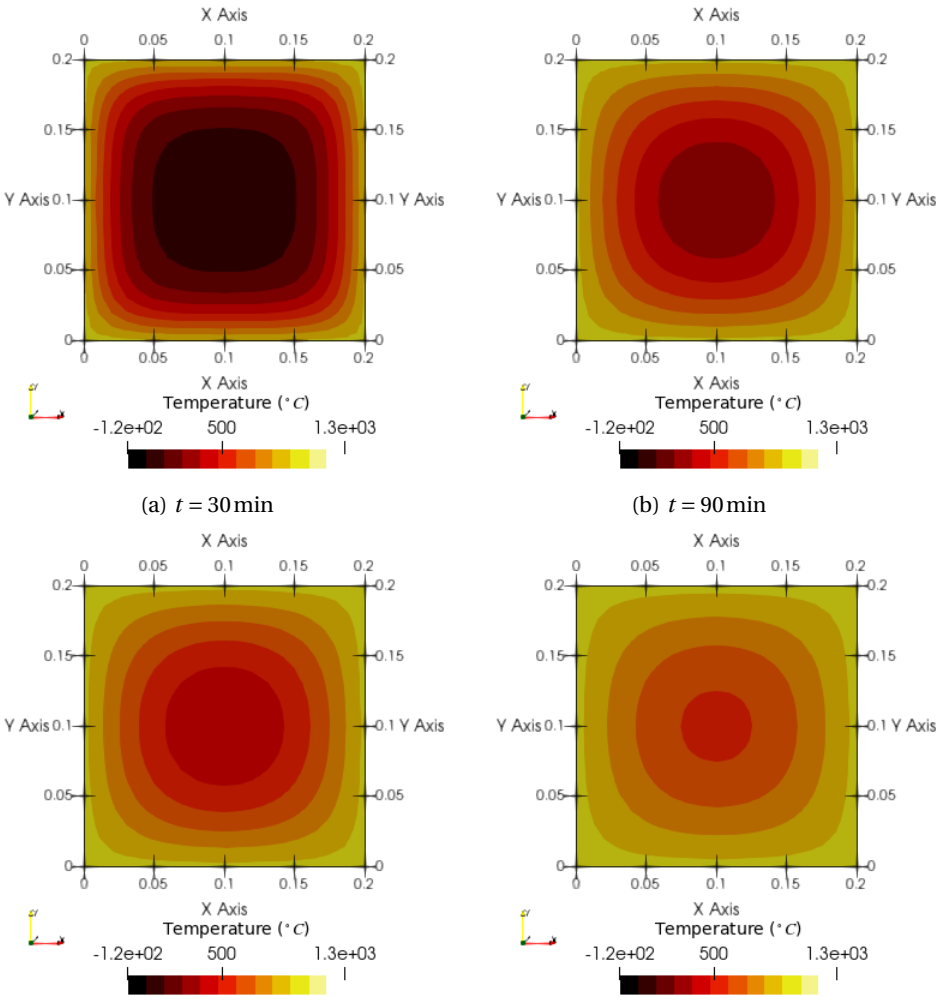
Table 5.4: Reference and computed values for T_0 concerning the validation example 2.

Time (min)	Reference value T_0 (°C)	Element Type	Computed value T_0' (°C)	Relative difference ε (%)
0	0.0	TRI3	0.0000	0.00
		TRI6	0.0000	0.00
		QUAD4	0.0000	0.00
		QUAD8	0.0000	0.00
30	36.9	TRI3	29.7312	1.94×10^{1}
		TRI6	33.5906	8.97
		QUAD4	30.4248	1.75×10^{1}
		QUAD8	33.8503	8.26
60	137.4	TRI3	130.0251	5.37
		TRI6	133.7875	2.63
		QUAD4	131.0145	4.65
		QUAD8	133.8905	2.55
90	244.6	TRI3	240.0627	1.85
		TRI6	242.8709	7.07×10^{-1}
		QUAD4	240.4040	1.72
		QUAD8	242.9500	6.75×10^{-1}
120	361.1	TRI3	362.2362	3.15×10^{-1}
		TRI6	363.4852	6.61×10^{-1}
		QUAD4	361.9427	2.33×10^{-1}
		QUAD8	363.5435	6.77×10^{-1}
150	466.2	TRI3	470.0065	8.16×10^{-1}
		TRI6	470.2503	8.69×10^{-1}
		QUAD4	469.3439	6.74×10^{-1}
		QUAD8	470.2947	8.78×10^{-1}
180	554.8	TRI3	560.5277	1.03
		TRI6	560.1557	9.65×10^{-1}
		QUAD4	559.6558	8.75×10^{-1}
		QUAD8	560.1907	9.72×10^{-1}



(a) (b)

Figure 5.5: Numerical results regarding the evolution of the temperature distribution for the validation example 2. (a) Temperature values at X as a function of time. (b) Relative error in percentage as function of time.



(a) $t = 30\,\text{min}$ (b) $t = 90\,\text{min}$

(c) $t = 120\,\text{min}$ (d) $t = 180\,\text{min}$

Figure 5.6: Numerical results for the validation example 2 using a TRI3 mesh.

5.3 Validation example 3 - The Standard NAFEMS Benchmarks: linear thermo-elastic tests - Two dimensional heat transfer with convection

5.3.1 Description

The geometry examined is a rectangular plate with width equal to 0.6 m and length equal to 1 m, as shown in Figure 5.7. A corresponding three-dimensional geometry is also considered with a thickness equal to 1 m. The boundary conditions considered are as follows: the left edge is assumed to be adiabatic. At the lower edge the temperature is prescribed to be 100 °C and along the upper and right edges there is heat transfer by convection and radiation. The heat convection coefficient h is equal to 70 $\mathrm{Wm^{-2}K^{-1}}$, and the ambient temperature is equal to 0 °C (see Equation 3.17). The initial temperature for the entire plate is 0 °C. The relevant properties of the material making up the plate are its conductivity k, equal to 52 $\mathrm{Wm^{-1}K^{-1}}$, its specific heat c_p, equal to 1 $\mathrm{Jkg^{-1}K^{-1}}$, and its density ρ, set equal to 1 $\mathrm{kgm^{-3}}$. Table 5.5 summarizes all the information regarding initial and boundary conditions, geometry and material properties. The expected temperature at E (see Figure 5.7) is 18.3 °C.

5.3.2 Results

The numerical solutions obtained using FEM are presented in Table 5.6 for two dimensions and in Table 5.7 for three-dimensions, as well as, the reference values and the corresponding relative difference. It can be seen that that agreement between the numerical results and the reference solutions is acceptable. It is below 1% for all elements employed, except for the TETRA4 element. There is no significant difference between the two integrators tested. Figure 5.8 shows the temperature distribution obtained using TRI3 and TETRA10 elements, which is reasonable given the description of the problem.

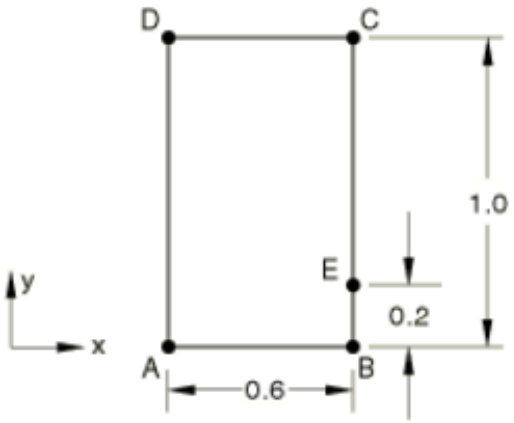


Figure 5.7: Geometry and boundary conditions considered in the validation example 3 (**?**).

Table 5.5: Material properties, and initial and boundary conditions for validation example 3.

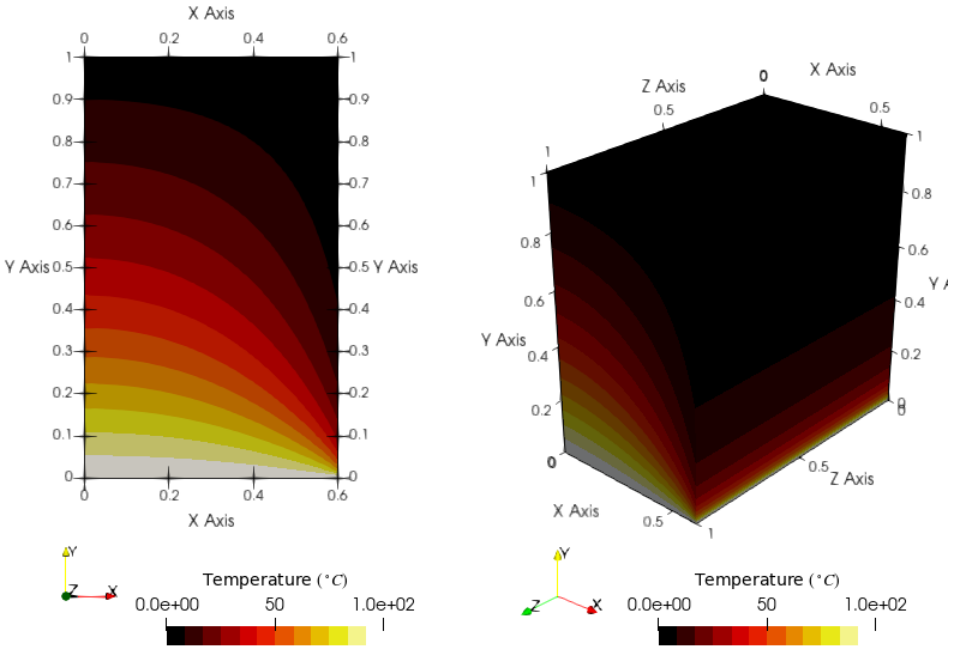
Material Properties			Effective value
Conductivity	k	($Wm^{-1}K^{-1}$)	52
Specific heat	c_p	($Jkg^{-1}K^{-1}$)	1
Density	ρ	(kg/m^3)	1
Boundary Conditions			
Dimension	h	(m)	1
Dimension	b	(m)	0.6
Thickness	t	(m)	1
Heat convection coefficient	h_c	($W/m^2/K$)	70
Initial Conditions			
Ambient temperature	T_∞	(°C)	100
Temperature in cross-section	T_0	(°C)	0
Reference value			
Temperature T_0 at point E		(°C)	

Table 5.6: Reference and computed values for T_0 concerning the validation example 3 in two-dimensions.

Element	Temperature T_0 at E °C	Relative difference ε (%)
Alpha integrator ($\rho = 1$)		
TRI3	18.1890	6.07×10^{-1}
TRI6	18.2553	2.44×10^{-1}
QUAD4	18.2286	3.90×10^{-1}
QUAD8	18.2532	2.56×10^{-1}
Quasi static integrator		
TRI3	18.1895	6.04×10^{-1}
TRI6	18.2548	2.47×10^{-1}
QUAD4	18.2281	3.93×10^{-1}
QUAD8	18.2536	2.54×10^{-1}

Table 5.7: Reference and computed values for T_0 concerning the validation example 3 in three-dimensions.

Element	Temperature T_0 at E °C	Relative difference ε (%)
Alpha integrator ($\rho = 1$)		
TETRA4	17.9501	1.91
TETRA10	18.2556	2.43×10^{-1}
Quasi static integrator		
TETRA4	17.9497	1.91
TETRA10	18.2548	2.47×10^{-1}



(a) (b)

Figure 5.8: Temperature distribution concercing the validation example 3: (a) in two-dimensions (TRI6) (b) in three-dimensions (TETRA4).

Chapter 6

Solution procedures for coupled fields

An overview of solution techniques for coupled fields is provided in this chapter. It comprises methods used to tackle a range of coupled field problems, including aeroelasticity, fluid-structure interaction, and thermomechanical coupling. Its purpose is to aid in the selection of solution techniques for thermoplastic problems that are accurate, stable, efficient in terms of memory and computational effort, and simple to implement and subsequently expand to other couplings, such as the ones found in electro-thermomechanical problems.

6.1 Context field elimination

Field elimination achieves the solution of a coupled problem by eliminating the variables of the first field and introducing them into the second field. This second field is then solved.

The decrease in the number of state variables is this procedure's key benefit. It results in less complex equation systems that should be simpler to solve. Other variables can also be chosen by the analyst so that they become the variables of interest. In this manner, it is not necessary to retrieve the variables that were removed. (**?**)

On the other hand **?** cite as disadvantages

- only possible for problems allowing explicit (and well-conditioned) variable eliminations;
- sparseness and symmetry attributes of matrices associated with the original coupled system can be adversely affected by the eliminations process; and
- available software modules for the isolated fields are not likely to be of much use for processing the reduced system.

The remainder of the chapter disregards these procedures, including in Section **??** where the comparison of the different schemes is discussed.

6.2 Monolithic

Monolithic algorithms solve the coupled nonlinear multi-physics system simultaneously. Implicit techniques are typically used to provide strong stability characteristics. Likewise, the Newton-Raphson technique is frequently used to solve the nonlinear residual equations. The effective solution of a large system of equations, including any potential nonlinearities or loss of symmetry, is a particular difficulty for monolithic algorithms. Even the units chosen can contribute to the ill-conditioning of the system matrix. Thus, a good preconditioning strategy is a key component of effective solvers for large-scale problems.

6.2.1 Numerical considerations

For the solution of a large system of equations, iterative methods are preferable to direct methods, in part, due to memory footprint considerations. The Newton-Krylov methods such as GMRES and the BiCGStab are among the most commonly used in multi-physics problems (**?**). However, their use does not suffice for an efficient and robust solution procedure for a multi-physics problem. In addition, the use of preconditioners alleviates the possible large condition numbers of the system matrix. There are several preconditioning techniques for the solution of large systems of equations, e.g., ILU preconditioners, domain decomposition, including multigrid approaches; multilevel recursive Schur complements preconditioners (see **?** and **?**).

? is concerned with the fully coupled solution of large-displacement fluid-structure interaction problems by Newton's method. They use block-triangular approximations of the Jacobian matrix, obtained by neglecting selected fluid-structure interaction blocks, and show that they provide suitable preconditioners for the solution of the linear systems with GMRES. A Schur complement approximation for the Navier-Stokes block and multigrid approximations for the solution of the computationally most expensive operations is the basis for the efficient approximate implementation of the preconditioners.

? propose a method based on a fully implicit, monolithic formulation of the problem in the arbitrary Lagrangian-Eulerian framework to solve the problem of fluid-structure interaction of an incompressible elastic object in laminar incompressible viscous flow They utilize the standard geometric multigrid approach based on a hierarchy of grids obtained by successive regular refinement of a given coarse mesh. The complete multigrid iteration is performed in the standard defect-correction setup with the V or F-type cycle.

? show how preconditioning techniques more sophisticated than diagonal preconditioning can be used in iterative solutions of the linear equation systems in fluid-structure interaction problems.

In **?**, the authors focus on the strong coupling fluid-structure interaction employing monolithic solution schemes. Therein, a Newton-Krylov method is applied to the monolithic set of nonlinear equations. They propose two preconditioners that apply algebraic multigrid techniques to the entire fluid-structure interaction system of equations. As the first option, the authors employ a standard block Gauss-Seidel approach, where approximate inverses of the individual field blocks are based on an algebraic multigrid hierarchy tailored to the type of the underlying physical problem. A monolithic coarsening scheme for the coupled system that uses prolongation and restriction projections constructed for the individual fields provides the basis for the second preconditioner. The resulting nonsymmetric monolithic algebraic multigrid

method involves coupling the fields on coarse approximations to the problem yielding significantly enhanced performance, claim the authors.

In the context of multi-physics problems, **?** propose a fully coupled algebraic multilevel preconditioner for Newton-Krylov solution methods. A set of multi-physics partial differential equation (PDE) applications attests its performance: a drift-diffusion approximation for semiconductor devices, a low Mach number formulation for the simulation of coupled flow, transport, and non-equilibrium chemical reactions, a low Mach number formulation for visco-resistive magnetohydrodynamics (MHD) systems. An aggressive-coarsening graph-partitioning of the non-zero block structure of the Jacobian matrix provides the basis for the algebraic multilevel preconditioner. Using a different approach **?** employ a new family of recursive block LU preconditioners to solve the thermally coupled induction less magnetohydrodynamics problem equations, which model the flow of an electrically charged fluid under the influence of an external electromagnetic field with thermal coupling.

? addresses a thermo-mechanically coupled problem of thermo-viscoelasticity at finite strains using a monolithic approach. The authors solve the system of nonlinear algebraic equations obtained from the spatial (FEM) and temporal (diagonally-implicit Runge-Kutta methods) discretization of the problem monolithically. They employ the Multilevel-Newton algorithm to obtain a high-order result in the space and the time domain. The numerical concept is applied to a constitutive model of finite strain thermo-viscoelasticity. **?** also employ in the context of thermo-viscoelasticity, the multilevel Newton algorithm to solve the system of algebraic equations describing the discretized problem.

? presents a monolithic solution scheme for thermo-structure interaction problem, using right preconditioning and a GMRES. The preconditioner "sub-problem" is solved using a Richardson iteration scheme and a relaxed block Gauss-Seidel method, which uncouples the mechanical and thermal problems. This procedure tackles each problem using an independent algebraic multigrid (AMG) preconditioner. **?** also consider the procedure just mentioned, as well as a preconditioner based on a semi-implicit method for pressure-linked equations, extended to deal with an arbitrary number of fields. This technique also results in uncoupled problems that can be solved with standard AMG. They also introduce a more sophisticated preconditioner that enforces the coupling at all AMG levels, unlike the other two techniques, which resolve the coupling only at the finest level. These techniques are applied successfully to three different coupled problems: thermo-structure interaction, fluid-structure interaction, and a complex model of the human lung.

? propose a hybrid interface preconditioner for the monolithic solution of surface-coupled problems. They combine physics-based block preconditioners with an additional additive Schwarz preconditioner, whose subdomains span across the interface on purpose. This approach is motivated by the error assessment of physics-based block preconditioners, revealing an accumulation of the error at the coupling surface, despite their overall efficiency.

6.2.2 Usage examples

Thermo-mechanical coupling In the following paragraph, a small overview of the literature is presented regarding the application of monolithic solvers to the thermo-mechanical coupled problem. **?** suggests a monolithic approach to the thermoelastic problem at small strains. The constitutive laws considered do not acknowledge the dependence of the mechanical properties on the temperature and are not deduced from a Helmholz energy function. **?** uses monolithic algorithms for the calculation of thin-walled structures using shell elements and an arc-length method for the TSI solution. While all coupling terms were considered, only a simplified mechanical dissipation was included where the hardening power was neglected (according to **?**). **?** present a thermoplasticity covariant formulation within the framework of the principal axis methodology, which the authors claim, leads to a very efficient numerical implementation. The paper contains several numerical simulations dealing with the fully coupled thermomechanical response at large viscoplastic strains, including strain localization and cyclic loading cases, to illustrate the performance of the proposed methodology. The authors consider the von Mises thermoplasticity yield criterion and strain energy depending on logarithmic stretches, a hardening variable, and temperature. A monolithic solver achieves the solution to the coupled problem, but no details about it are given. **?** proposes a volume-coupled TSI model based on the finite element method for the structural and thermal field. Various temperature-dependent, isotropic, elastic, and elastoplastic material models for small and finite strains are employed, incorporating the effect of the highly elevated temperatures predominating in rocket nozzles, the practical application focused on in the Ph.D. thesis. The author considers both monolithic and partitioned coupling algorithms to solve fully coupled thermomechanical systems. Regarding the former, a novel monolithic Newton-Krylov scheme with problem-specific block Gauss-Seidel preconditioner and algebraic multigrid methods is introduced. Concerning the latter, loosely and strongly coupled partitioned schemes are examined, possibly including acceleration techniques, as, e.g., the Aitken Δ^2 method. **?** and **?** both present monolithic approaches, based on the multilevel Newton method, for the solution of the thermo-mechanical problem. In both contributions, thermo-visco-plastic materials are successfully analyzed. Recently, **?** have presented a finite strain thermo-mechanically coupled two-surface damage-plasticity theory. The authors obtain the solution for the three coupled fields, displacement, nonlocal damage variable, and temperature, employing an implicit and monolithic solution scheme.

The thermo-mechanical coupling has also been studied in the more specific domain of contact mechanics. **?** present one of the earliest contributions in this direction. They propose a FEM formulation of frictionless contact, accounting for full thermo-elastic coupling. The penalty method is used to enforce the non-penetration conditions. Another contribution, **?**, advances a standard mortar discretization with Lagrange multipliers to solve the small strain thermo-elasticity problem. The authors consider the heat equation coupled to linear mechanics through a thermal expansion term in their formulation. The solution approach is based on a preconditioned Jacobian-free Newton Krylov solution method, and the use case under analysis is a light water reactor nuclear fuel rod. **?** investigate thermomechanical mortar contact algorithms and their application to NURBS-based Isogeometric Analysis in the context of nonlinear elasticity. Mortar methods are applied to both the mechanical and thermal fields to model frictional contact, the energy transfer between the surfaces,

and frictional heating. A monolithic approach is pursued in solving the nonlinear algebraic equations found after the discretization in time and space. In the Ph.D. thesis by the same first author, **?**, this approach is further pursued in multi-field contact problems. More recently, **??** tackles the numerical treatment of contact problems considering inelastic deformation and thermomechanical coupling. It accounts for plastic spin, visco-plasticity, and thermo-plastic coupling, as well as temperature-dependent material parameters. The authors also opt for a monolithic solver, although no further details are supplied. See also, in the context of contact mechanics, **?????** and **?**.

Others In the context of fluid-structure interaction, the monolithic approach seems to be more widely used than in thermo-mechanically coupled problems. A few contributions in this domain using a monolith approach are **???????????** and **?**. The use of a monolithic approach can also be found in the domain of saturated soils (e.g., **?**, **?**, **?**, **?**). Monolithic solvers are also used in the context of magnetohydrodynamics (e.g., **?** and **?**).

6.3 Partitioned

The following section presents the partitioned time-stepping algorithms. For a detailed comparison with the monolithic approach and between themselves, see Section 6.4.

A field partition is a field-by-field decomposition of the space discretization. Partitioning may be algebraic or differential. In algebraic partitioning, the complete coupled system is spatially discretized first and then decomposed. In differential partitioning, the decomposition is done first, and each field is then discretized separately. Differential partitioning often leads to non-matched meshes, as is typical of fluid-structure interaction. Algebraic partitioning was initially developed for matched meshes and substructuring (**?**).

The earliest contributions regarding the partitioned treatment of coupled systems emerged in the mid 1970s, involving structure-structure interactions and fluid-structure interactions (see e.g. **?**, **?**, **?**, **?** and **?**).

Given a complex system, there are usually many ways of partitioning it into subsystems or fields. **?** provide a very pragmatic and helpful criterion to select the fields to be considered. According to their definition, a field is characterized by computational considerations. It is a segment of the overall problem for which a separable software module is either available or readily prepared if the interaction terms are suppressed. As such, a partitioned approach to the solution of multi-physics problems employs field analyzers specific to each field separately stepped in time. The coupling between the fields is achieved through proper communication between the individual components using prediction, substitution, and synchronization techniques.

Before moving on, it may be helpful to clear up the difference between partitioned schemes, staggered schemes, operator splits, and fractional-step methods. The first is probably the most general term and includes the others. Its definition has already been given. A staggered scheme is a term most often used for the partitioned schemes where the solution concerning each field is sequential and obtained only once per time step as in the loosely coupled schemes to be introduced. However, it may also include the strongly coupled schemes, as well. An operator split is obtained through the decomposition of the fully coupled problem into subproblems. The structure of

the problem is the same, as well as the unknowns considered. The only difference between the subproblems is the physical effects considered. The equation terms concerning each physical effect must be divided exclusively and exhaustively between the subproblems. Finally, according to **?** staggered algorithms for coupled problems can be viewed as fractional steps methods, in the sense of **?**, arising from an operator split of the coupled problem of evolution.

6.3.1 Operator splits

The most common operator splits into thermomechanical problems are the isothermic and adiabatic split.

Isothermic The isothermic split is perhaps the most straightforward and natural approach, as noted by **?**, one of the earliest contributions on the topic. The scheme achieves the solution of the thermo-mechanical problem, first solving the mechanical problem at a constant temperature, then a purely thermal phase is considered at a fixed configuration.

Adiabatic The adiabatic split is proposed in **?**. It consists of a first phase where constant entropy is enforced and a second phase of purely thermal conduction with a fixed reference. In terms of implementation complexity, it is comparable to the isothermal split. This is possible because the constant entropy phase can be cast as a mechanical phase, where the stiffness and the external force are adjusted as a function of an intermediate temperature. This temperature is computed considering the strong form of the temperature evolution equation to retain the computational efficiency of the isothermal split, despite the momentum equation being enforced in its weak form. The advantage of this split is that when used in a staggered scheme, it is unconditionally stable (see Section 6.4).

6.3.2 Loosely vs. Strongly coupled schemes

According to **?** there are several basic techniques associated with partitioned schemes (see Figure 6.1). These are

- preditction;
- substitution;
- interfield iteration;
- full step correction;
- lockstep advancing;
- midpoint correction;
- subcycling;
- augmentation.

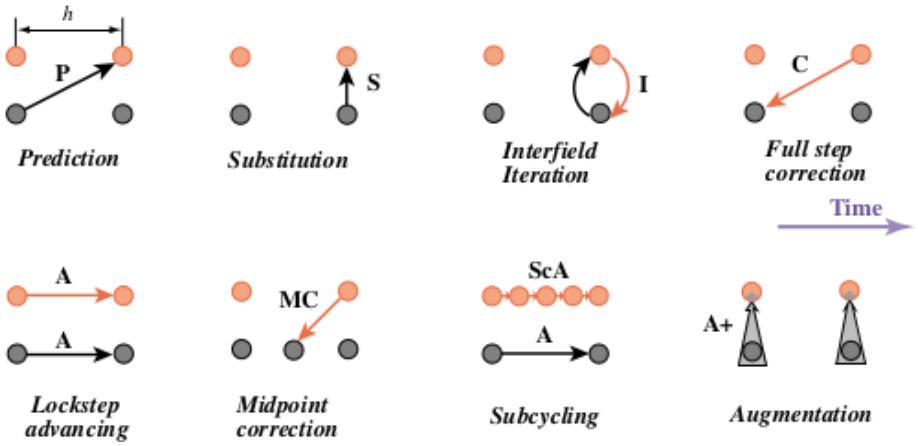


Figure 6.1: Devices of partitioned analysis time-stepping (**?**).

Inter-field iterations are the primary criterion distinguishing loosely or one-way staggered coupled schemes and strongly or iterative staggered coupled schemes. In the loosely coupled schemes, the integration algorithm proceeds sequentially, solving the problem in each field only once per time step. On the other hand, for strongly coupled schemes, inter-field iterations are present, such that the problems are solved multiple times at the same time instant. This inner loop is repeated until a given tolerance is reached for the unknowns in each field.

The remainder of the techniques listed will be mentioned and explained in the discussion below.

6.3.3 Loosely coupled

The solution for the fully coupled problem is found in loosely coupled schemes by solving each field sequentially. For the thermomechanical problem, the two available schemes are the isothermal split (see e.g. **?**, **?**) and the adiabatic split (see e.g. **?** and **?**), as mentioned above.

According to **?**, in linear problems, the first concern with partitioning is the degradation of time-stepping stability. After the analyst has ensured stability, an accuracy analysis of the method should be performed. In strongly nonlinear problems, such as fluid flow, stability and accuracy tend to be intertwined since numerical stability is harder to define. As such, they are usually considered together in method design. The expectation is for a method that operates well at a reasonable timestep.

? present a detailed explanation about how to design from scratch a loosely coupled time-stepping algorithm applicable to linear systems of equations. It includes implementation details, such as the choice of the predictor formula, and the design steps, from the formulation of the original field equations and temporal discretization to the stability and accuracy analysis. Other contributions focused mainly on linear systems are **?**, **?** and **?**,

Because the loosely coupled schemes are explicit, they are also often only conditionally stable. The isothermic split is such an example (**?**). On the other hand, the adiabatic split proposed in **?** is unconditionally stable, despite being explicit. **?** also propose a stable staggered scheme, achieved through semi-algebraic

augmentation, which, however, is limited to linearized thermoelasticity. In the context of coupled flow and geomechanics, **?** show that when the mechanical problem is solved first, the drained split combined with a backward Euler discretization is conditionally stable, and the undrained split is unconditionally stable when combined with midpoint rule. When instead the flow problem is solved first, **?** show that the fixed-stress split is conditionally and the fixed-strain split is unconditionally stable for appropriate choices of the generalized midpoint rule.

Moreover, in the domain of fluid-structure interactions, it can be shown that staggered methods are inherently non-conservative. As time progresses in the simulation, these schemes introduce parasitic energy at the boundary, which contributes to their poor numerical stability (**?**). A further problem appears when solving these coupled physical problems, the so-called artificial added-mass effect, which leads to instability. It manifests itself when a slender structure and fluid have similar densities, and the latter is modeled as an incompressible fluid (**??**). It can even be shown that for every sequentially staggered scheme and spatial discretization of a problem, a mass ratio between the fluid and structural mass density can be found at which the coupled system becomes unstable (**?**).

Despite this, some contributions detail strategies allowing for the unconditional stability of these schemes. As part of the development loop of commercial tire designs, **?** tackles the problem of tire hydroplaning. The author presents a robust explicit coupling scheme that relies on rigorous control of the energy artificially introduced at the interface by the staggering process through a dynamic adaptation of the coupling time step size. Regarding the artificial added-mass effect, **?** demonstrates that even for fluid-structure applications with strong added mass effects, a carefully designed staggered and sub-iteration-free time-integrator can achieve numerical stability and robustness concerning the slenderness of the structure, as long as the fluid is justifiably modeled as a compressible medium.

Another technique available to improve the stability of loosely coupled schemes is (algebraic) augmentation. It rests on the injection of one of the coupled equations into the other, after discretization in space, to 'soften' the system, either by reducing the large eigenvalues of the uncoupled stiff equation or by introducing some damping into it. Some examples of this approach include **?** and **?**.

Yet another technique to ensure stability in the context of fluid-structure interactions is presented in **?**. Stability is achieved employing a semi-implicit coupling scheme, splitting the added-mass, viscous effects, and geometrical/convective nonlinearities, through a Chorin-Temam projection scheme within the fluid.

Regarding accuracy, the loosely coupled schemes do not necessarily inherit the accuracy order of the schemes used in the integration of the separate fields, often being just of the first order in time (**?**). However, some contributions detail approaches that are second-order time-accurate. In the context of thermo-elasticity, **?** show that a double-pass approach using the adiabatic split yields such a second-order accurate time-stepping algorithm. A few approaches yield similar results in the domain of fluid-structure interaction (see **?**, **?** and **?**). In any case, whatever the theoretical convergence order of the loosely coupled method, at a given time instant, the fully coupled discretized equations of the problem will never be exactly satisfied by the solutions found. There is a lag between the fields considered, e.g., the mechanical and thermal fields in a thermomechanical problem. In the context of strong coupling, this lag can be conceived as a numerical evaluation error. Solving approximately the exact (i.e., aggregated) equations can be reinterpreted as exactly solving a set of approximate

(i.e., segregated) equations. Thus, one can construe loosely-coupled methods as solving a set of segregated equations instead of aggregated equations. Accordingly, the incurred numerical evaluation error can be reinterpreted as a discretization error. Loosely-coupled methods, therefore, satisfy conservation only in an asymptotic sense, i.e., for vanishing mesh width; this is a basic consistency requirement (**?**).

Prediction techniques can improve the order of the numerical evaluation error incurred by loosely-coupled partitioned methods. For the sake of explanation, consider the thermo-mechanical problem being solved using the isothermic split. When using predictors, instead of integrating the mechanical equations based on the structure's temperature in the previous time instant, a prediction can be used for the temperature of the structure boundary in the current time instant. Such predictions are generally based on an extrapolation of the solution from the previous time step. Prediction techniques improve the solution accuracy and stability of loosely-coupled methods (**????**).

Another technique available to improve the accuracy of the loosely coupled methods is subcycling. It involves solving each field's problems using different time steps since the fields present in a multi-physics problem often have different time scales. In the context of aeroelasticity, **?** claims that it can offer substantial computational advantages, including savings in the simulation CPU time because the structural field will be advanced fewer times. **?** and **?** also argue for this technique along the same lines.

Usage examples The loosely coupled scheme has been used in the context of thermoelasticity (**??????**), thermo-plasticity (**??????**) and thermo-viscoplasticity (**??**).

For examples in aeroelasticity see e.g., **??** and **?**, and in fluid-structure interaction more broadly see e.g. **?** and **?** Other applications include fluid-soil interaction analysis (**???**).

6.3.4 Strongly coupled

In the strongly coupled scheme, inter-field iterations are performed until a given tolerance for the unknowns of each field is reached. They converge to the solution of the monolithic scheme and are thus able to satisfy discrete versions of the coupled problem exactly (**??**). In principle, regarding thermomechanics, either the isothermal or the adiabatic slit can be used, but there seems to be no example of the latter. In contrast to the staggered schemes, there is no problem of conditional stability, but the scheme may converge very slowly or not at all. As an example coming from fluid-structure interactions, it has been shown that the number of coupling iterations increases when the time step decreases or when the structure becomes more flexible (**?**). This can place a severe restriction on the use of these schemes. Several acceleration techniques are available in the literature to speed up convergence.

A straightforward way to improve the convergence behavior of the strongly coupled schemes is using predictors, in contrast to the values found in the last step. Thus, the initial guesses can be improved using well-chosen predictors **?**. Along these lines, **?** employ polynomial prediction methods, and **?** use a line extrapolation method to improve the first guess of the unknown and thus decrease the number of iterations needed to achieve convergence.

Another approach that is well established for series acceleration is the Aitken delta-squared process. It uses previously computed values to obtain more accurate

estimates for the unknown. **?** is an early contribution detailing this low-memory convergence acceleration scheme. In the context of thermomechanics, **?**, **?** and **?** use this technique, with the last authors also employing a quasi-Newton least squares method. Some examples of contributions in the domain of fluid-structure interactions taking advantage of this approach are **?**, **?** and **?**. The last authors also introduce a vector extrapolation approach that includes more than three previous values of the iteration scheme in the improved estimate.

The strongly coupled approach lends itself to an interpretation as a nonlinear block Jacobi or Gauss-Seidel scheme, whose convergence is conditional and at most linear (**??**). **?** provides an in-depth analysis of block Jacobi and Gauss-Seidel schemes applied to coupled problems, including considerations regarding efficiency, complexity, and parallelization. **??** suggests a block-Newton method instead, with the Jacobian of the system being approximate by a finite difference method. Under some assumptions on the subsystem solvers, this approach converges quadratically. **?** propose a solution method based on the conjugation of sub-iterations via a Newton-Krylov method, which confines the GMRES acceleration to the interface degrees-of-freedom. The latter renders storage requirements for the Krylov space and computational cost of the least-squares problem low. The nesting of Newton and GMRES iterations lends itself to the reuse of Krylov vectors in subsequent linear system solutions. **?** claims that the approach proposed by the last authors should not be regarded as a Newton-based solver but as a Krylov-based vector extrapolation scheme

One can also improve the convergence speed of the strongly coupled scheme using reduced-order models to produce a more accurate first guess and thus decrease the number of iterations needed for the method to converge. **?** presents a technique that uses the Jacobian from reduced-order models that are built up during the coupling iterations. The reduced-order model is built for each step and approximates an arbitrary interface displacement fitting a linear regression to the previous displacement-stress points. **?** follows the same technique, coupling it with an Aitken delta-squared process.

? proposes a manifold mapping technique to decrease the number of sub-iterations of a high-fidelity fluid-structure interaction model. The idea is to perform many sub-iterations with a low-fidelity model instead of the high-fidelity flow and structure models.

Usage examples Regarding the use of strongly coupled schemes in the context of thermo-mechanics, there are a few contributions. **?** present results concerning thermo-elasticity at finite strains, **?** concerning thermo-viscoelasticity, **?** includes results on thermo-elasticity and thermo-elasto-plasticity. In field of fluid-structure interaction, a few examples of the use of strongly coupled schemes are **?**, **?** and **?**. Including more than two fields, **?** tackles electro-thermo-mechanical problems, as does **?**, which also considers radiative heat transfer. In **?**, the strongly coupled scheme is used to solve coupled hygro-thermo-mechanical problems in photovoltaic laminates.

6.4 Comparison of solution techniques

According to **?**, the desirable properties of a time-stepping algorithm for solving coupled problems are:

- enjoys unconditional stability;
- is highly accurate;
- is easy to implement;
- is not memory intensive;
- requires low CPU time;
- satisfies software modularity constraints.

In the following, the time-stepping schemes presented above are compared with these criteria in mind. The application in view is thermomechanics.

Stability Regarding stability, the loosely coupled using an isothermal split is conditionally stable (**?**). Despite this, the limitation is not significant for metal plasticity, according to **?**. However, examples where the scheme diverges, can be found in **?**. In this last contribution, the adiabatic split is introduced and shown to be unconditionally stable in the context of thermo-elasticity. **?** show that these properties extend to thermo-plasticity. The strongly coupled schemes are unconditionally stable because no critical time step leads to numerical instabilities in the results. Despite this, the inner loop of the scheme may converge slowly or not at all (**?**). It depends on the spectral radius of the matrices involved (**?**). There are, however, acceleration techniques that can mitigate this problem, including predictors and Aitken Δ^2 methods (see Section **??**). **?** presents a numerical example concerning an internal pressurized thick-walled cylinder, whose material is viscoplastic, for which the strongly coupled scheme employed diverged, despite the use of an Aitken method. On the other hand, the monolithic scheme, as long as appropriately preconditioned, is unconditionally stable (**?**).

Accuracy Regarding accuracy, the solution found from the loosely coupled method will never exactly satisfy the fully coupled discretized equations of the problem. There will be a time lag between the thermal and the mechanical field. Loosely-coupled methods, therefore, satisfy conservation only in an asymptotic sense, i.e., for vanishing mesh width (**?**). As long as it does not diverge, the monolithic and strongly coupled satisfy the coupled discretized equations exactly.

Ease of implementation The partitioned schemes are much easier to implement as most of them can work with the field analyzers as black boxes, concerning themselves only with communication between the solvers, initial guesses, and acceleration schemes using previously computed values. The monolithic scheme requires the computation of the full stiffness matrix, including the mixed terms and appropriate preconditioning that varies widely with the specific multi-physics problem to be solved.

Memory requirements When it comes to memory requirements, the partitioned schemes often require only the diagonal blocks of the stiffness matrix found in the linearization process. Previous values also need to be saved from one iteration to the next, increasing the memory cost for some acceleration techniques. In contrast, the fully coupled monolithic scheme requires the full stiffness matrix of the coupled problem.

CPU time According to **?**, solving a fluid-structure interaction problem with the same accuracy using a loosely and strongly coupled scheme, the latter is more efficient than the former. For the same total number of iterations, the difference in the accuracy reached ranges from one to three orders of magnitude. These results run counter to a claim in **?**. However, this is not supported by any numerical results from the last authors. In the numerical examples presented in **?**, the monolithic solver is in most cases faster than a strongly coupled scheme employing an Aitken method for problems in thermomechanics. The differences range from 120% to 140% in favor of the monolithic scheme. Supporting evidence for these conclusions can also be found in **?**. The authors report CPU time ratios between the strongly coupled and monolithic approaches, ranging from 0.635 to 3.75 on the magnitude of the coupling.

Software modularity The partitioned approaches can take full advantage of software, including closed source commercial solvers. There is little to no software reuse for the monolithic approach, save for routines that solve linear systems and the like.

Conclusions Lastly, it may be helpful to reproduce the recommendations given in **?** regarding the choice between partitioned and monolithic approaches. According to the authors, the circumstances that favor the partitioned approach for tackling a coupled problem are a research environment with few delivery constraints, access to existing software, localized interaction effects (e.g., surface versus volume), and widespread spatial/temporal component characteristics. The opposite circumstances: commercial environment, rigid deliverable timetable, massive software development resources, global interaction effects, and comparable length/time scales favors a monolithic approach.

Putting it all together, the most appropriate choice for the present use case is the strongly coupled schemes with appropriate acceleration techniques. They can take advantage of already existing software, provide accurate results that agree with a monolithic approach, are not memory intensive, are easy to implement, and with the use of convergence acceleration techniques, are competitive from the computational efficiency standpoint. The only drawback seems to be the possibility of divergence in the inner loop, stalling the progress of the simulation.

Table 6.1: Summary of the comparison between the FFT-Galerkin method.

	Partitioned schemes		Monolithic
	Loosely coupled	Strongly coupled	
Stability	Isothermic split: conditionally stable Adiabatic split: unconditionally stable	unconditionally stable*	unconditionally stable
Accuracy	Coupled discretized equations not satisfied exactly	Coupled discretized equations satisfied	Coupled discretized equations satisfied
Ease of implementation	Only communication between field analyzers stricly needed		Full coupling needed: • Computation of mixed terms of the Jacobian • Preconditioning needed
Memory requirements	Only diagonal blocks of the full stiffness matrix needed		Full stiffness matrix needed
Software modularity constraints	Full software modularity		Poor or no software modularity

* The inner loop of the strongly coupled scheme may converge very slowly or even diverge.

Page intentionally left blank.

Chapter 7

Implicit solution methods for coupled fields

This chapter presents the most common strongly coupled/implicit methods employed to solve coupled field problems. This presentation seeks to provide a literature overview of the available approaches.

7.1 Equations to be solved

For the sake of clarity, the discretized equations of the thermo-mechanical problem at the next time instant, $n+1$e are recovered here

$$\mathbf{M}\ddot{\mathbf{u}}_{n+1} + \mathbf{f}_u^{\text{int}}(\theta_{n+1}, \mathbf{u}_{n+1}) - \mathbf{f}_{u,n+1}^{\text{ext}} = \mathbf{0}, \tag{7.1}$$

$$\mathbf{C}\dot{\theta}_{n+1} + \mathbf{K}\theta_{n+1} + \mathbf{f}_\theta^{\text{int}}(\theta_{n+1}, \mathbf{u}_{n+1}) - \mathbf{f}_{\theta,n+1}^{\text{ext}} = \mathbf{0}. \tag{7.2}$$

The complete definition of the material incremental discretized thermo-mechanical initial boundary value problem can be found in Chapter 4.

As only partitioned approaches are considered, the thermal and mechanical problems are solved separately, i.e., Equation (7.1) is solved considering a fixed temperature, and Equation (7.2) is solved assuming a fixed configuration. To ease the discussion, consider the existence of two functions $\mathcal{U}_{n+1}$ and $\mathcal{T}_{n+1}$ that represent these solution procedures at timestep $n+1$. These so-called mechanical and thermal solvers satisfy

$$\mathcal{U}: \mathscr{K}_{\theta,n+1} \to \mathscr{K}_{u,n+1}, \quad \mathbf{u} = \mathcal{U}_{n+1}(\theta), \tag{7.3}$$

$$\mathcal{T}: \mathscr{K}_{u,n+1} \to \mathscr{K}_{\theta,n+1}, \quad \theta = \mathcal{T}_{n+1}(\mathbf{u}). \tag{7.4}$$

See Chapter 4 for detailed information on them. In the following, the subscripts on the solvers will be dropped to avoid clutter.

The goal now is to consider functions, built from $\mathcal{U}$ and $\mathcal{T}$, whose roots are also the solutions to the thermo-mechanical problem (Equations (7.1) and (7.2)). Several examples can be provided. The most appropriate for the current use case are presented in what follows. They can be found in **?** in the context of fluid-structure interaction (FSI).

Consider the residues defined as,

$$\mathcal{R}_{\mathrm{J}}\colon \mathcal{K}_{u,n+1}\times\mathcal{K}_{\theta,n+1}\to K_{u,n+1}\times\mathcal{K}_{\theta,n+1},\quad \mathcal{R}_{\mathrm{J}}(\mathbf{u},\theta)=\left\{\begin{array}{c}\mathbf{u}-\mathcal{U}(\boldsymbol{\theta})\\ \theta-\mathcal{T}(\mathbf{u})\end{array}\right\}, \tag{7.5}$$

and

$$\mathcal{R}_{\mathrm{GS}}\colon \mathcal{K}_{\theta,n+1}\to\mathcal{K}_{\theta,n+1},\quad \mathcal{R}_{\mathrm{GS}}(\theta)=\theta-\mathcal{T}\circ\mathcal{U}(\theta), \tag{7.6}$$

or

$$\mathcal{R}^{*}_{\mathrm{GS}}\colon \mathcal{K}_{u,n+1}\to\mathcal{K}_{u,n+1},\quad \mathcal{R}^{*}_{\mathrm{GS}}(\mathbf{u})=\mathbf{u}-\mathcal{U}\circ\mathcal{T}(\mathbf{u}), \tag{7.7}$$

where the subscript "J" stands for Jacobi and the subscript "GS" for Gauss-Seidel. The reason for this choice of subscripts is made clear in Section 7.3.1.

Since the methods described below for the solution of nonlinear systems of equations apply to both functions $\mathcal{R}_{\mathrm{J}}$ and $\mathcal{R}_{\mathrm{GS}}$, a general function denoted as $\mathcal{R}$, whose variable is $\mathbf{x}$, is considered instead. As already stated, the solution for the thermo-mechanical problem (Equations (7.1) and (7.2)) can be abstracted as the solution of

$$\mathcal{R}(\mathbf{x})=0. \tag{7.8}$$

To obtain simpler expressions in what follows, consider also the function

$$\mathcal{S}(\mathbf{x})=\mathbf{x}-\mathcal{R}(\mathbf{x}), \tag{7.9}$$

whose fixed point is the solution to the nonlinear system of equation in Equation (7.8).

7.2 A classification scheme for iterative methods

Most methods available for the solution of systems of nonlinear equations, such as the one in Equation (7.8), are iterative methods. They can be more precisely defined lettting $\mathbf{x}^{k},\mathbf{x}^{k-1},\ldots$, whose superscripts correspond to the loop of the iteration method, be approximants to $\mathbf{x}_{n+1}$, whose subscript concerns the timestep

To better understand the landscape of available methods to solve nonlinear systems of equations, the iteration functions are classified according to the information they require following the classification scheme by **?**. Let $\mathbf{x}^{k+1}$ be determined uniquely by information obtained at $\mathbf{x}^{k},\mathbf{x}^{k-1},\ldots$, including the derivatives of any order of $\mathcal{R}$. Let the function that maps $\mathbf{x}^{k},\mathbf{x}^{k-1},\ldots$ into $\mathbf{x}^{k+1}$ be called ϕ. Thus

$$\mathbf{x}^{k+1}=\phi\left(\mathbf{x}^{k},\mathcal{R}(\mathbf{x}^{k}),J_{\mathcal{R}}(\mathbf{x}^{k}),\ldots\right), \tag{7.10}$$

where ϕ is called an iteration function, and $J_{\mathcal{R}}$ is the Jacobian of $\mathcal{R}$. To prevent clutering $\mathbf{x}^{k}$ will stand for its value as well as for the values of $\mathcal{R}(\mathbf{x}^{k})$, $J_{\mathcal{R}}(\mathbf{x}^{k})$ and further derivatives of higher order. Then ϕ is called a *one-point iteration function*. Most iteration functions that have been used for root-finding are one-point iteration functions. The most commonly known examples are the fixed point schemes and Newton's iteration method.

Next, let $\mathbf{x}^{k+1}$ be determined by new information at $\mathbf{x}^{k}$ and reused information at $\mathbf{x}^{k-1},\ldots$ Thus

$$\mathbf{x}^{k+1} = \phi\left(\mathbf{x}^k; \mathbf{x}^{k-1}, \ldots\right). \tag{7.11}$$

Then ϕ is called a *one-point iteration function with memory*. The semicolon in Equation (7.11) separates the point at which new data are used from the points at which old data are reused. The secant iteration function is the best-known example of a one-point iteration function with memory.

Let $\mathbf{x}^{k+1}$ be determined by new information at $\mathbf{x}^k, \omega_1\left(\mathbf{x}^k\right), \ldots, \omega_i\left(\mathbf{x}^k\right)$, $i \geq 1$, where ω_i denote operations on $\mathbf{x}^k$. No old information is reused. Thus

$$\mathbf{x}^{k+1} = \phi\left[\mathbf{x}^k, \omega_1\left(\mathbf{x}^k\right), \ldots, \omega_i\left(\mathbf{x}^k\right)\right]. \tag{7.12}$$

Then ϕ is called a *multipoint iterative function*. Such methods include the Aitken-Steffson method.

Finally, let $\mathbf{z}_j$ represent the quantities $\mathbf{x}^j, \omega_1\left(\mathbf{x}^j\right), \ldots, \omega_i\left(\mathbf{x}^j\right)$, $i \geq 1$. Let

$$\mathbf{x}^{k+1} = \phi\left(\mathbf{z}^k; \mathbf{z}^{k-1}, \ldots\right). \tag{7.13}$$

Then ϕ is called a *multipoint iterative function with memory*. The semicolon in Equation (7.13) separates the points at which new data are used from the points at which old data are reused.

In the present work, the criteria used for the choice of the iterative method used fit roughly into the ones provided by **?** for problems in the context the electronic structure problems. They are

1. The dimensionality of the problem is large.
2. $\mathcal{R}(\mathbf{x})$ is continuously differentiable, but the analytic form of its derivative is not readily available, or it is costly to compute.
3. The cost of evaluating $\mathcal{R}(\mathbf{x})$ is computationally high.
4. The problem is noisy. In other words, the evaluated function values of $\mathcal{R}(\mathbf{x})$ usually contain errors.

Thus, the methods chosen must minimize the number of calls to $\mathcal{R}$, as it is expensive to compute. The amount of information saved from previous iterations must also be judiciously chosen as the problem's dimensionality is large, leading to memory limitations. Finally, the analytical form of the derivative $\mathcal{R}$ is also not available. Thus methods that use it must be discarded.

7.2.1 Predictor

Iterative procedures are considered to solve the thermo-mechanical problem at a given timestep $n+1$. As the first value approximating $\mathbf{x}_{n+1}$, one can employ the converged value of the previous timestep, $\mathbf{x}_n$. However, a very efficient way to increase the chances of stability and reduce computation time is to predict the optimal initial values at the beginning of every time step (**???**). The prediction of the new solution by polynomial

extrapolation is based on the converged solution of the last two or three timesteps. This method is based on polynomial vector extrapolation, which is relatively easy to implement, and the extra computational input is negligible.

The maximum polynomial under consideration is of the order two, i.e., the new solution is extrapolated from the results from the last three time steps. The predictors $\mathbf{x}^*$ for the order $p=1$ and $p=2$ polynomials read:

$$p=1: \quad \mathbf{x}^*_{n+1} = 2\mathbf{x}_n - \mathbf{x}_{n-1}, \tag{7.14}$$

$$p=2: \quad \mathbf{x}^*_{n+1} = 3\mathbf{x}_n - 3\mathbf{x}_{n-1} + \mathbf{x}_{n-2}. \tag{7.15}$$

7.2.2 Global Approaches

Following **?**, the terms "global," as in "global method" or "globally convergent algorithm," are here used to denote a method that is designed to converge to a local minimizer of a nonlinear functional or some solution of a system of nonlinear equations, from almost any starting point. The methods presented in this chapter do not qualify as global methods since if the initial trial is not close enough to the solution, they will not converge. There are, however, approaches to mitigate this problem. The ideas presented below apply with particular relevance to the Newton method (see Section 7.3.2) and related procedures. Their exposition follows **?** where more details can be found.

Consider that the iterative solution method determines $\Delta\mathbf{x}^k$ in

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \Delta\mathbf{x}^k. \tag{7.16}$$

The two global approaches here considered both come into action after $\Delta\mathbf{x}^k$ has been computed by some appropriate method (see from Section 7.3 on). At this point, one decides whether to accept the step $\Delta\mathbf{x}^k$ or to choose $\mathbf{x}^{k+1}$ by a global strategy.

A solution to the system of equations (7.8) clearly also satisfies

$$r(\mathbf{x}) = 0, \quad \text{where } r \equiv \frac{1}{2}\|\mathcal{R}\|_2^2 \colon \mathbb{R}^n \to \mathbb{R}, \tag{7.17}$$

so the problem can be regarded as an unconstrained minimization problem, with caveat that local minimizers of r may not be the solution to the system of equations (7.8).

The basic idea of a global method for unconstrained minimization is geometrically obvious: take steps that lead "downhill" for the function r. More precisely, one chooses a direction $\mathbf{p}$ from the current point $\mathbf{x}^k$ in which r decreases initially, and a new point $\mathbf{x}^{k+1}$ in this direction from $\mathbf{x}^k$ such that $r(\mathbf{x}^{k-1}) < r(\mathbf{x}^k)$. Such a direction is called a descent direction.

An important question to ask is, "What is a descent direction for problem (7.17)?". It is any direction $\mathbf{p}$ for which $\nabla r(\mathbf{x}^k)^T\mathbf{p} < 0$, where

$$\nabla r\left(\mathbf{x}^k\right) = J_{\mathcal{R}}\left(\mathbf{x}^k\right)^T \mathcal{R}\left(\mathbf{x}^k\right), \tag{7.18}$$

where $J_{\mathcal{R}}(\mathbf{x}^k)$ is the Jacobian matrix of $\mathcal{R}$ at $\mathbf{x}^k$. Therefore, the steepest-descent direction for (7.17) is along $-J_{\mathcal{R}}(\mathbf{x}^k)^T\mathcal{R}(\mathbf{x}^k)$.

The Newton step for the update equation (7.16) is (see Section 7.3.2)

$$\Delta \mathbf{x}_N^k = -J_{\mathcal{R}}\left(\mathbf{x}^k\right)^{-1} \mathcal{R}\left(\mathbf{x}^k\right), \tag{7.19}$$

and it is a descent direction, since

$$\nabla r\left(\mathbf{x}^k\right)^T \Delta \mathbf{x}_N^k = -\mathcal{R}\left(\mathbf{x}^k\right)^T J_{\mathcal{R}}\left(\mathbf{x}^k\right) J_{\mathcal{R}}\left(\mathbf{x}^k\right)^{-1} \mathcal{R}\left(\mathbf{x}^k\right) = -\mathcal{R}\left(\mathbf{x}^k\right)^T \mathcal{R}\left(\mathbf{x}^k\right) < 0 \tag{7.20}$$

as long as $\mathcal{R}\left(\mathbf{x}^k\right) \neq \mathbf{0}$. Hence, the appropriateness of these methods to the Newton method and related methods.

Since the Newton step yields a root of

$$M^k\left(\mathbf{x}^k + \Delta \mathbf{x}^k\right) = \mathcal{R}\left(\mathbf{x}^k\right) + J_{\mathcal{R}}\left(\mathbf{x}^k\right) \Delta \mathbf{x}^k, \tag{7.21}$$

it also goes to a minimum of the quadratic function

$$\begin{aligned} \hat{m}^k\left(\mathbf{x}^k + \Delta \mathbf{x}^k\right) &\equiv \frac{1}{2} M^k\left(\mathbf{x}^k + \Delta \mathbf{x}^k\right)^T M^k\left(\mathbf{x}^k + \Delta \mathbf{x}^k\right) \\ &= \frac{1}{2} \mathcal{R}\left(\mathbf{x}^k\right)^T \mathcal{R}\left(\mathbf{x}^k\right) + \left(J_{\mathcal{R}}\left(\mathbf{x}^k\right)^T \mathcal{R}\left(\mathbf{x}^k\right)\right)^T \Delta \mathbf{x}^k \\ &\quad + \frac{1}{2} \Delta \mathbf{x}^{k\,T} \left(J_{\mathcal{R}}\left(\mathbf{x}^k\right)^T J_{\mathcal{R}}\left(\mathbf{x}^k\right)\right) \Delta \mathbf{x}^k, \end{aligned} \tag{7.22}$$

because $\hat{m}^k\left(\mathbf{x}^k + \Delta \mathbf{x}^k\right) \geq 0$ for all $\Delta \mathbf{x}^k$ and $\hat{m}^k\left(\mathbf{x}^k + \Delta \mathbf{x}_N^k\right) = 0$. Therefore, $\Delta \mathbf{x}_N^k$ is a descent direction for $\hat{m}^k$, and since the gradients at $\mathbf{x}^k$ of $\hat{m}^k$ and r are the same, it is also a descent direction for r.

The above development motivates how the global methods to be described are applied, i.e., they are applied to the quadratic model $\hat{m}^k(\mathbf{x})$. Since $\nabla^2 \hat{m}^k(\mathbf{x}^k) = J_{\mathcal{R}}(\mathbf{x}^k)^T J_{\mathcal{R}}(\mathbf{x}^k)$, this model is positive definite as long as $J_{\mathcal{R}}(\mathbf{x}^k)$ is nonsingular, which is consistent with the fact that $\mathbf{x}^k + \Delta \mathbf{x}_N^k$ is the unique root of $M^k(\mathbf{x})$ and thus the unique minimizer of $\hat{m}^k(\mathbf{x})$ in this case. Thus, the model $\hat{m}^k(\mathbf{x})$ has the attractive properties that its minimizer is the Newton point for the original problem, and that all its descent directions are descent directions for $r(\mathbf{x})$ because $\nabla \hat{m}^k(\mathbf{x}^k) = \nabla r(\mathbf{x}^k)$. Therefore methods based on this model, by going downhill and trying to minimize $\hat{m}^k(\mathbf{x})$, will combine Newton's method for nonlinear equations with global methods for an associated minimization problem.

If the Jacobian of $\mathcal{R}$ is not available and its estimate is of poor quality, the global procedure may be compromised (**?**). However, these procedures may be unnecessary in the present use case since the initial trial is probably close enough to the solution, even without accounting for the improvements coming from more carefully chosen initial shots through predictors (see Section 7.2.1). **?** also employ a simple restarting procedure instead of a global convergence strategy. If in two consecutive values of $\mathcal{R}$, $\mathcal{R}_{\text{old}}$ and $\mathcal{R}_{\text{new}}$, $\|\mathcal{R}_{\text{new}}\|$ is much larger than $\|\mathcal{R}_{\text{old}}\|$, the solution procedure is restarted, with the new initial trial values corresponding to $\mathcal{R}_{\text{old}}$. They suggest r between 0.1 and 0.3, with $\|\mathcal{R}_{\text{old}}\| < r\|\mathcal{R}_{\text{new}}\|$ leading to a restart. In their opinion, global approaches such as those suggested below are too expensive when the evaluation of $\mathcal{R}$ is also costly to compute.

Line search The line search approach is based on the traditional idea of backtracking along the Newton direction if a complete Newton step is unsatisfactory. More precisely given a descent direction $\mathbf{p}^k$, a step in that direction is taken as

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \lambda^k \mathbf{p}^k, \tag{7.23}$$

for some $\lambda^k > 0$ that makes $\mathbf{x}^{k+1}$ an acceptable iterate. The common procedure is to first try $\lambda_k = 1$ and only if this fails backtrack in a systematic way along the direction defined by that step. See **?** for a full discussion on the choice of λ_k.

Trust region algorithms The trust region algorithm is based on estimating the region in which the local model, underlying Newton's method, can be trusted to represent the function adequately and taking a step to approximately minimize the model in this region. It drops the assumption that the step must be in the Newton direction. $\hat{m}^k(\mathbf{x}^k + \Delta\mathbf{x}^k)$ is approximately minimized subject to $\|\mathbf{x}^k\|_2 \leq \delta^k$. If $\delta^k \geq \left\| J_{\mathcal{R}}(\mathbf{x}^k)^{-1}\mathcal{R}(\mathbf{x}^k) \right\|_2$, then the step attempted is the Newton step. Otherwise, for the locally constrained optimal step, it is

$$\Delta\mathbf{x}^k = -\left(J_{\mathcal{R}}\left(\mathbf{x}^k\right)^T J_{\mathcal{R}}\left(\mathbf{x}^k\right) + \mu^k \mathbf{I} \right)^{-1} J_{\mathcal{R}}\left(\mathbf{x}^k\right)^T \mathcal{R}\left(\mathbf{x}^k\right), \tag{7.24}$$

for μ^k such that $\|\Delta\mathbf{x}^k\|_2 \cong \delta^k$. For the details on the choice of δ^k see **?**.

7.2.3 Convergence criteria

For an iterative method to be useful, there must be reasonable criteria to determine its convergence. The iteration residual is defined as

$$\mathbf{r}^k = \mathcal{R}(\mathbf{x}^k), \tag{7.25}$$

and if it is equal to zero then $\mathbf{x}$ is the solution to the system of nonlinear equations, i.e.,

$$\mathbf{r} = \mathcal{R}(\mathbf{x}) = \mathbf{0}, \tag{7.26}$$

and hence, a reasonable convergence measure for the iteration procedure.

The discrete l^2-norm can be used to obtain a scalar representative of the vectorial residual $\mathbf{r}^k = \left(r^{k,1}, \ldots, r^{k,n_{\text{unknown}}} \right)^T$ as

$$\left\| \mathbf{r}^k \right\|_{L^2} = \sqrt{\sum_i \left(r^{k,i} \right)^2}. \tag{7.27}$$

Directly using (7.27) yields an absolute convergence criterion

$$\left\| \boldsymbol{r}^k \right\|_{l^2} < \epsilon_{\text{abs}}. \tag{7.28}$$

with $\epsilon_{\text{abs}} > 0$ as an absolute convergence tolerance, with convergence being achieved when the above condition is satisfied. However, since the absolute value of the $r^{k,i}$'s can change by orders of magnitude during one simulation, an absolute measure is not appropriate in all situations. A relative measure solves this problem by setting the residual in relation with the current coupling iterate values as

$$\frac{\left\| \mathbf{r}^k \right\|_{l^2}}{\left\| \mathbf{x}^k \right\|_{l^2}} < \epsilon_{\text{rel}}. \tag{7.29}$$

A relative convergence measure can fail to work correctly when the coupling iterate values are close to zero, and rounding errors occur. Thus, a combination of absolute and relative measures, where the absolute measure takes care of close to zero cases, and the relative handles all other cases, is often a good choice.

7.3 One-point iteration function

7.3.1 Fixed-point approaches

The application of the fixed-point method to obtain the roots of $\mathcal{R}$ yields

$$\mathbf{x}^{k+1} = \mathcal{S}(\mathbf{x}^k) = \mathbf{x}^k - \mathcal{R}(\mathbf{x}^k). \tag{7.30}$$

See Figure 7.1 for its geometric interpretation in one dimension.

If the particular functions defined on Equations (7.5) and (7.6) are used, one finds the two basic Schwarz procedures commonly employed in strongly coupled solution procedures. They are the additive or block Jacobi and the parallel Scharwz or Gauss-Seidel procedures. The names originate from domain decomposition, and justify the subscripts employed in Equations (7.5) and (7.6).

7.3.1.1 Block Jacobi or Schwarz additive

Applying the fixed-point approach to $\mathcal{R}_\mathrm{J}$ (Equation (7.5)), yields

$$\left\{\mathbf{u}^{k+1}, \theta^{k+1}\right\}^T = \mathcal{S}_\mathrm{J}(\mathbf{u}^k, \theta^k) \tag{7.31}$$

$$= \left\{\mathbf{u}^k, \theta^k\right\}^T - \mathcal{R}_\mathrm{J}(\mathbf{u}^k, \theta^k), \tag{7.32}$$

It is the same as solving both the mechanical (Equation (7.1)) and the thermal problem (Equation (7.2)) in parallel. Such a procedure is said to be Schwarz additive or block Jacobi, refering to the similarities with the procedure for the solution of linear systems of equations with the same name i.e.,

$$\mathbf{u}^{k+1} = \mathcal{U}(\theta^k), \tag{7.33}$$

$$\theta^{k+1} = \mathcal{T}(\mathbf{u}^k). \tag{7.34}$$

Box 7.1 shows the pseudo-code for the block Jacobi approach.

7.3.1.2 Block Gauss-Seidel or Schwarz multiplicative

Applying the fixed-point approach to $\mathcal{R}_\mathrm{GS}$ (Equation (7.5)), yields

$$\theta^{k+1} = \mathcal{S}_\mathrm{GS}(\theta^k) = \theta^k - \mathcal{R}_\mathrm{GS}(\theta^k). \tag{7.35}$$

Thus, the fields are solved sequentially, where the output of the first solver is the input of the second solver. This the solution procedure is said to be Scharwz multiplicative or block Gauss-Seidel.

$$\mathbf{u}^{k+1} = \mathcal{U}(\theta^k), \tag{7.36}$$

$$\theta^{k+1} = \mathcal{T}(\mathbf{u}^{k+1}). \tag{7.37}$$

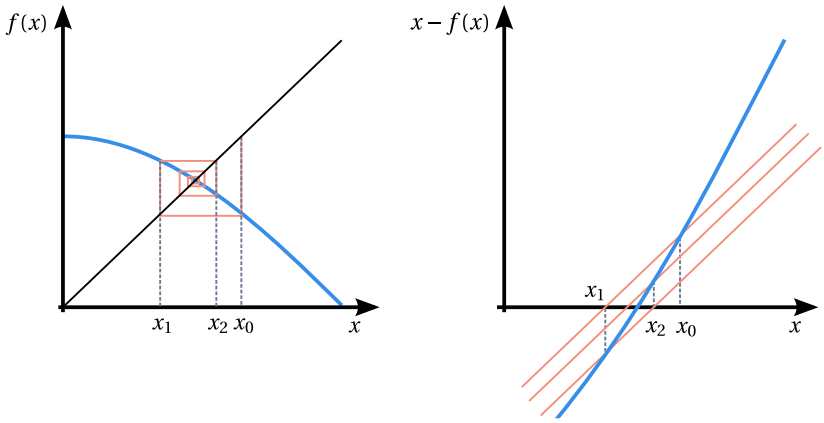


Figure 7.1: Geometric interpretation of the fixed-point iteration method in one dimension. The fixed-point of f is sought, which is equivalent to the root of $x - f(x)$.

One of the fields must be chosen as the first, and this may be crucial for the stability and convergence rate of the approach (**?**). Here, the focus is on the sequence coinciding with the isothermic split, i.e., first, the mechanical problem is solved at a fixed temperature. Then the thermal problem is solved at a fixed configuration.

Box 7.2 shows the pseudo-code for the block Gauss-Seidel approach.

7.3.2 Newton's method

The Newton-Raphson or Newton scheme is a very popular iterative solution procedure for nonlinear systems of equations, which under appropriate conditions converges quadratically (**??**). It can be applied to Equation (7.8) yielding

$$J_{\mathcal{R}}(\mathbf{x}^k)\Delta\mathbf{x}^k = -\mathcal{R}(\mathbf{x}^k), \tag{7.38}$$

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \Delta\mathbf{x}^k. \tag{7.39}$$

See Figure 7.2 for its geometric interpretation in one dimension.

In particular, using $\mathcal{R}_\mathrm{J}$, a few simplifications can be obtained. To ease the explanation, consider, a thermal residual operator $\mathcal{R}_u(\mathbf{u},\theta)$ and a mechanical residual operator $\mathcal{R}_\theta(\mathbf{u},\theta)$ defined to be the first and second components in the definition of $\mathcal{R}_\mathrm{J}$ (Equation (7.6)). Written in full

$$\mathcal{R}_u(\mathbf{u},\theta) = \mathbf{u} - \mathcal{U}(\theta) = 0, \tag{7.40}$$

$$\mathcal{R}_\theta(\mathbf{u},\theta) = \theta - \mathcal{T}(\mathbf{u}) = 0, \tag{7.41}$$

From this, a block Newton iteration can be written as

Box 7.1: Additive Schwarz procedure, also called block Jacobi, for one timestep.

(i) $\mathbf{u}^0 = \mathbf{u}_n$

(ii) $\theta^0 = \theta_n$

(iii) Set fixed-point counter to zero: $k = 0$

(iv) Enter the fixed-point loop

(1) Solve the mechanical problem at fixed temeperature θ^k: $\mathbf{u}^{k+1} = \mathcal{U}(\theta^k)$

(2) Solve the thermal problem at a fixed configuration $\mathbf{u}^k$: $\theta^{k+1} = \mathcal{T}(\mathbf{u}^k)$

(3) If the desired accuracy has not been reached, update $k = k+1$ and go to step (1).

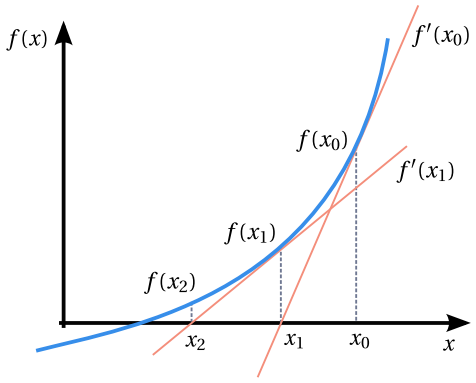


Figure 7.2: Geometric interpretation of the Newton method in one dimension for an example function f, whose derivative is denoted by f'.

$$\begin{bmatrix} J_{\mathcal{R}_u}\left(\mathbf{u}^k, \theta^k\right) \\ J_{\mathcal{R}_\theta}\left(\mathbf{u}^k, \theta^k\right) \end{bmatrix} \left\{ \begin{array}{c} \Delta\mathbf{u}^k \\ \Delta\theta^k \end{array} \right\} = - \left\{ \begin{array}{c} \mathcal{R}_u\left(\mathbf{u}^k, \boldsymbol{\theta}^k\right) \\ \mathcal{R}_\theta\left(\mathbf{u}^k, \boldsymbol{\theta}^k\right) \end{array} \right\}, \tag{7.42}$$

and the update of the iteration variables reads

$$\left\{ \begin{array}{c} \mathbf{u}^{k+1} \\ \boldsymbol{\theta}^{k+1} \end{array} \right\} = \left\{ \begin{array}{c} \mathbf{u}^k \\ \boldsymbol{\theta}^k \end{array} \right\} + \left\{ \begin{array}{c} \Delta\mathbf{u}^k \\ \Delta\boldsymbol{\theta}^k \end{array} \right\}. \tag{7.43}$$

The system of equations in Equation (7.42) can be further simplified following **?** considering the definitions of the mechanical and thermal residuals and taking their

Box 7.2: Multiplicative Schwarz procedure, also called block Gauss-Seidel, for one timestep.

(i) $\mathbf{u}^0 = \mathbf{u}_n$

(ii) $\theta^0 = \theta_n$

(iii) Set fixed-point counter to zero: $k = 0$

(iv) Enter the fixed-point loop

(1) Solve the mechanical problem at fixed temeperature θ^k: $\mathbf{u}^{k+1} = \mathcal{U}(\theta^k)$

(2) Solve the thermal problem at a fixed configuration $\mathbf{u}^{k+1}$: $\theta^{k+1} = \mathcal{T}(\mathbf{u}^{k+1})$

(3) If the desired accuracy has not been reached, update $k = k+1$ and go to step (1).

derivatives. It yields

$$\begin{bmatrix} \mathbf{I} & -J_{\mathcal{U}}\left(\theta^k\right) \\ -J_{\mathcal{T}}\left(\mathbf{u}^k\right) & \mathbf{I} \end{bmatrix} \begin{Bmatrix} \Delta\mathbf{u}^k \\ \Delta\theta^k \end{Bmatrix} = - \begin{Bmatrix} \mathcal{R}_u\left(\mathbf{u}^k, \boldsymbol{\theta}^k\right) \\ \mathcal{R}_\theta\left(\mathbf{u}^k, \boldsymbol{\theta}^k\right) \end{Bmatrix}, \tag{7.44}$$

Solving for $\Delta\mathbf{u}^k$ and $\Delta\theta^k$, one finds

$$\left(\mathbf{I} + J_{\mathcal{U}}(\theta^k) J_{\mathcal{T}}(\mathbf{u}^k)\right) \Delta\mathbf{u}^k = -\mathcal{R}_u(\mathbf{u}^k, \theta^k) + J_{\mathcal{U}}(\theta^k)\mathcal{R}_\theta(\mathbf{u}^k, \theta^k), \tag{7.45}$$

$$\left(\mathbf{I} + J_{\mathcal{T}}(\mathbf{u}^k) J_{\mathcal{U}}(\theta^k)\right) \Delta\theta^k = -\mathcal{R}_\theta(\theta^k, \mathbf{u}^k) + J_\theta(\mathbf{u}^k)\mathcal{R}_u(\mathbf{u}^k, \theta^k). \tag{7.46}$$

Thus, the Jacobians now needed are $J_{\mathcal{U}}$ and $J_{\mathcal{T}}$. See Section 7.4.2 for the practical application of this.

Every iteration of the Newton scheme involves at least one invocation of the thermal and mechanical solvers when computing $\mathcal{R}\left(\mathbf{u}^k\right)$ or both $\mathcal{R}_u\left(\mathbf{u}^k, \boldsymbol{\theta}^k\right)$ and $\mathcal{R}_\theta\left(\mathbf{u}^k, \boldsymbol{\theta}^k\right)$. The critical point for black-box equation coupling is how to obtain the derivative information in the Jacobi matrices. In different ways, some of the methods presented next find approximations for the required Jacobian times vector products.

7.3.3 Constant Underrelaxation

One of the most straightforward ways to stabilize an iterative method is to use constant underrelaxation (**?**). The relaxation is performed as follows

$$\mathbf{x}^{k+1} = (1-\omega)\mathbf{x}^k + \omega\left(\mathbf{x}^k - \mathcal{R}(\mathbf{x}^k)\right) = \mathbf{x}^k - \omega\mathcal{R}(\mathbf{x}^k), \tag{7.47}$$

where ω is the relaxation factor chosen in the range $0 < \omega < 1$, which corresponds to an underrelaxation, to achieve a stabilizing effect.

Applying to Equation (7.5)

$$\left\{ \begin{array}{c} \mathbf{u}^{k+1} \\ \theta^{k+1} \end{array} \right\} = (1-\omega) \left\{ \begin{array}{c} \mathbf{u}^{k} \\ \theta^{k} \end{array} \right\} + \omega \left\{ \begin{array}{c} \mathcal{U}(\theta^k) \\ \mathcal{T}(\mathbf{u}^k) \end{array} \right\} \tag{7.48}$$

Applying to Equation (7.6)

$$\theta^{k+1} = (1-\omega)\theta^k + \omega \mathcal{T} \circ \mathcal{U}(\theta^k). \tag{7.49}$$

Constant underrelaxation works well if ω is close to 1 but leads to a slow convergence if ω has to be chosen close to 0. Thus, the constant underrelaxation method creates unmanageable computational costs for severe instabilities. Overrelaxation can also be considered, keeping in mind that for $\omega > 2$ convergence is lost. The optimal ω is not necessary the largest stable one (**?**) and has to be set empirically. In what follows, alternative methods are discussed to decrease the number of iterations necessary while maintaining stability.

Box 7.3: Constant underrelaxation applied to the block Gauss-Seidel scheme.

(i) $\theta^0 = \theta^p_{n+1}$

(ii) Set fixed-point counter to zero: $k = 0$

(iii) Enter the fixed-point loop

  (1) Solve the mechanical problem at fixed temeperature θ^k: $\mathbf{u}^{k+1} = \mathcal{U}(\theta^k)$

  (2) Solve the thermal problem at a fixed configuration $\mathbf{u}^{k+1}$: $\theta^{k+1} = \mathcal{T}(\mathbf{u}^{k+1})$

  (3) Compute θ^{k+1} using constant relaxation (Equation (7.47))

  (4) If the desired accuracy has not been reached, update $k = k+1$ and go to step (1).

7.4 One-point iteration function with memory

7.4.1 Aitken relaxation

The so-called Aitken Δ^2 relaxation method was introduced by **?** as a modified Aitken Δ^2 that does not require the computation of the function twice per iteration as in the original method. It has been widely used in the context of FSI (**??????**). It has also been used in the context of thermo-mechanics by **?**.

In the one-dimensional case, this method resembles the secant method applied to the fixed point problem, which can be used to solve nonlinear equations without differentiation. Calling it an Aitken method is perhaps a misnomer since, in the Aitken-Steffensen method, the function values are computed twice per iteration (see Section 7.5.3). It is more closely related to secant methods, reusing values from previous iterations. This version of Aitken's Δ^2 method provides a dynamic under

relaxation, which can be used to improve the convergence/stability properties of the coupling algorithm.

Assume that f is the function whose fixed point is sought. The linear interpolation between two points already known of the function, $(a, f(a))$ and $(b, f(b)$ is

$$y = \frac{f(b) - f(a)}{b - a}(x - a) + f(a). \tag{7.50}$$

The fixed point of this approximation is

$$c = \frac{f(b) - f(a)}{b - a}(c - a) + f(a). \tag{7.51}$$

Thus, after rearranging,

$$c = \frac{af(b) - bf(a)}{\big(a - f(a)\big) - \big(b - f(b)\big)} \tag{7.52}$$

This can be rewritten as

$$c = \big(1 - \omega_b\big)\, b + \omega_b f(b) \quad \text{with } \omega_b = \frac{a - b}{\big(a - f(a)\big) - \big(b - f(b)\big)} \tag{7.53}$$

Anticipating the next iteration step,

$$d = \big(1 - \omega_c\big)\, c + \omega_c f(c) \quad \text{with } \omega_c = \frac{c - b}{\big(b - f(b)\big) - \big(c - f(c)\big)} \tag{7.54}$$

a convenient expression for updating the relaxation factor may be found, i.e.

$$\omega_c = -\omega_b \frac{f(b) - b}{(c - f(c)) - (f(b) - b)}. \tag{7.55}$$

See Figure 7.3 for its geometric interpretation in one dimension.

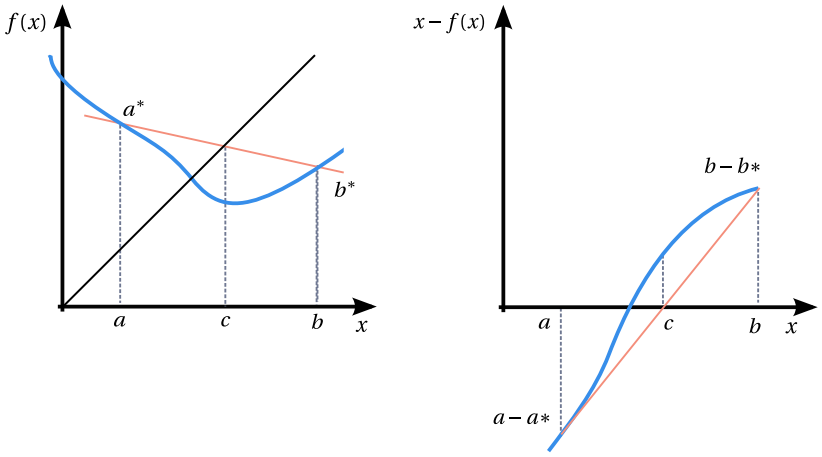


Figure 7.3: Geometric interpretation of the Aitken relaxation in one dimension for an example function f and corresponding interpreation as the secant method.

Now, for the vector case, the next step is to work out the solution to the current iteration from the outcome of the previous iteration $\mathbf{x}^k$ plus a new increment $\Delta\mathbf{x}^k$

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \Delta\mathbf{x}^k. \tag{7.56}$$

The increment reads

$$\Delta \mathbf{x}^k = \omega^k \left(\mathcal{S}(\mathbf{x}^{(k)}) - \mathbf{x}^{(k)} \right) = -\omega^k \mathcal{R}(\mathbf{x}^k). \tag{7.57}$$

with ω^k being the relaxation coefficient. This coefficient is updated in every iteration cycle as a function of two previous residuals

$$\omega^k = -\omega^{k-1} \frac{\left(\mathbf{r}^{(k)} - \mathbf{r}^{(k-1)} \right)^{\mathrm{T}} \mathbf{r}^{(k-1)}}{\left(\mathbf{r}^{(k)} - \mathbf{r}^{(k-1)} \right)^2}. \tag{7.58}$$

Comparing with Equations (7.38) and (7.39), ω^k can be, in a sense, regarded as an approximation to the inverse of the Jacobian. Dynamic relaxation is also easy to implement, and the additional computational input is acceptable since only inner vector products must be performed. The dynamical relaxation coefficient is restricted to the range $(0,2)$ because employing relaxation with a relaxation coefficient outside this range leads to loss of convergence (**?**). See Box 7.4 for the pseudocode.

Box 7.4: Aitken relaxation for one timestep.

(i) Set nonlinear counter to zero: $k = 0$

(ii) $\mathbf{x}^k = \mathbf{x}^p_{n+1}$

(iii) Enter the nonlinear loop

(1) Compute $\mathcal{R}(\mathbf{x}^k)$, which implies the solution of the mechanical and the thermal problems, $\mathcal{U}$ and $\mathcal{T}$, respectively.

(2) if $k = 0$:

- Compute $\mathbf{x}^{k+1}$ using constant relaxation (Equation (7.47))

(3) else:

- Compute $\mathbf{x}^{k+1}$ using Aitken relaxation relaxation (Equations (7.57) and (7.58))
- Save the current residue $\mathbf{r}^k = \mathcal{R}^k$.

(4) If the desired accuracy has not been reached, update $k = k+1$ and go to step (1).

7.4.2 Multi-secant methods

The following exposition follows closely **?**. In quasi-Newton methods the Jacobian is updated in each iteration using a rank-one update. Standard quasi-Newton methods require the updated J_{k+1} to satisfy the following secant condition

$$J^{k+1}_{\mathcal{R}} \Delta \mathbf{x}^k = \Delta \mathcal{R}^k, \tag{7.59}$$

where $\Delta\mathcal{R}^k \equiv \mathcal{R}\left(\mathbf{x}^{k+1}\right) - \mathcal{R}\left(\mathbf{x}^k\right)$. Furthermore, another common requirement is the following so-called no-change condition

$$J_{\mathcal{R}}^{k+1}\mathbf{q} = J_{\mathcal{R}}^{k}\mathbf{q} \quad \forall \mathbf{q} \text{ such that } \mathbf{q}^{\mathrm{T}}\Delta\mathbf{x}^k = 0, \tag{7.60}$$

which stipulates that there be no new information from $J_{\mathcal{R}}^k$ to $J_{\mathcal{R}}^{k+1}$ along any direction $\mathbf{q}$ orthogonal to $\Delta\mathbf{x}^k$.

? developed a method satisfying both secant condition (Equation (7.59)) and the no-change condition (Equation (7.60)), arriving at the update formula

$$J_{\mathcal{R}}^{k+1} = J_{\mathcal{R}}^{k} + \left(\Delta\mathcal{R}^k - J_{\mathcal{R}}^{k}\Delta\mathbf{x}^k\right)\frac{\Delta\mathbf{x}^{k\mathrm{T}}}{\Delta\mathbf{x}^{k\mathrm{T}}\Delta\mathbf{x}^k}. \tag{7.61}$$

Matrix $J_{\mathcal{R}}^{k+1}$ in Equation (7.61) is the unique matrix satisfying both conditions (7.59) and (7.60). The Broyden update can also be obtained by minimizing $E\left(J_{\mathcal{R}}^{k+1}\right) = \left\|J_{\mathcal{R}}^{k+1} - J_{\mathcal{R}}^{k}\right\|_F^2$ with respect to terms of $J_{\mathcal{R}}^{k+1}$, subject to the secant condition (7.59).

It may seem at first that Broyden's first method can be expensive since computing the quasi-Newton step $\Delta\mathbf{x}^k$ requires solving a linear system at each iteration. However, note that, typically, the approximate Jacobian is a small rank modification of a diagonal matrix (or a matrix that is easy to invert); hence, the cost to obtain this solution is not too high as long as the number of steps is not too large.

An alternative is Broyden's second method that approximates the inverse Jacobian instead of the Jacobian itself. $G_{\mathcal{R}}^k$ is used to denote the estimated inverse Jacobian at the k th iteration. The secant condition (Equation (7.59)) now reads

$$G_{\mathcal{R}}^{k+1}\Delta\mathcal{R}^k = \Delta\mathbf{x}^k \tag{7.62}$$

By minimizing $E\left(G_{\mathcal{R}}^{k+1}\right) = \left\|G_{\mathcal{R}}^{k+1} - G_{\mathcal{R}}^{k}\right\|_F^2$ with respect to $G_{\mathcal{R}}^{k+1}$ subject to Equation (7.62), the following update formula is found for the inverse Jacobian

$$G_{\mathcal{R}}^{k+1} = G_{\mathcal{R}}^{k} + \left(\Delta\mathbf{x}_k - G_{\mathcal{R}}^{k}\Delta\mathcal{R}^k\right)\frac{\Delta\mathcal{R}^{k\mathrm{T}}}{\Delta\mathcal{R}^{k\mathrm{T}}\Delta\mathcal{R}^k} \tag{7.63}$$

which is also the only update satisfying both the secant condition (Equation (7.62)) and the no-change condition for the inverse Jacobian

$$G_{\mathcal{R}}^{k}\mathbf{q} = G_{\mathcal{R}}^{k+1}\mathbf{q} \quad \forall \mathbf{q} \text{ such that } \mathbf{q}^{\mathrm{T}}\Delta\mathcal{R}^k = 0. \tag{7.64}$$

The update formula in Equation (7.61) can also be obtained in terms of $G_{\mathcal{R}}^k \equiv {J_{\mathcal{R}}^k}^{-1}$ by applying the Sherman-Morrison formula

$$G_{\mathcal{R}}^{k+1} = G_{\mathcal{R}}^{k} + \left(\Delta\mathbf{x}^k - G_{\mathcal{R}}^{k}\Delta\mathcal{R}^k\right)\frac{\Delta\mathbf{x}^{k\mathrm{T}}G_{\mathcal{R}}^{k}}{\Delta\mathbf{x}^{k\mathrm{T}}G_{\mathcal{R}}^{k}\Delta\mathcal{R}^k} \tag{7.65}$$

This shows, as was explained earlier, that to solve the Jacobian system associated with Broyden's first approach can be reduced to a set of update operations that are not more

costly than those required by the second update. Note, however, that the above formula requires the inverse of the initial Jacobian.

From Equation (7.61) and Equation (7.63) it is possible to define Broyden's family of updates, in which an update formula takes the general form

$$G_{\mathcal{R}}^{k+1} = G_{\mathcal{R}}^{k} + \left(\Delta \mathbf{x}^{k} - G_{\mathcal{R}}^{k} \Delta \mathcal{R}^{k}\right) \mathbf{v}_{k}^{\mathrm{T}} \tag{7.66}$$

where $\mathbf{v}_k^{\mathrm{T}} \Delta \mathcal{R}^k = 1$ so that the secant condition (7.59) holds. Note that the secant condition (7.62) is equivalent to condition (7.59). Some authors called Broyden's first method Broyden's good update and Broyden's second method as Broyden's bad update. These are two particular members of Broyden's family.

7.4.2.1 Generalized Broyden

The multi-secant methods provide an approximation to the Jacobian in Equation (7.38) or Equation (7.42) using information from previous iterations. A generalized Broyden's method with a flexible rank update on the inverse Jacobian, satisfying a set of m secant equations

$$G_{\mathcal{R}}^{k} \Delta \mathcal{R}^{i} = \Delta \mathbf{x}^{i} \quad \text{for } i = k-m, \ldots, k-1 \tag{7.67}$$

where it is assumed $\Delta \mathcal{R}^{k-m}, \ldots, \Delta \mathcal{R}^{k-1}$ are linearly independent and $m \leqslant n$ can also be described. Aggregating Equations (7.67) in matrix form, they can be rewriten it as

$$G_{\mathcal{R}}^{k} \mathscr{R}^{k} = \mathscr{X}^{k}. \tag{7.68}$$

where

$$\mathscr{R}^{k} = \left[\Delta \mathcal{R}^{k-m} \cdots \Delta \mathcal{R}^{k-1}\right], \quad \mathscr{X}^{k} = \left[\Delta \mathbf{x}^{k-m} \cdots \Delta \mathbf{x}^{k-1}\right] \in \mathbb{R}^{n \times m} \tag{7.69}$$

The no-change condition corresponding to (7.60) is

$$\left(G_{\mathcal{R}}^{k} - G_{\mathcal{R}}^{k-m}\right) \mathbf{q} = 0 \tag{7.70}$$

for all $\mathbf{q}$ orthogonal to the subspace spanned by $\Delta \mathcal{R}^{k-m}, \ldots, \Delta \mathcal{R}^{k-1}$, the columns of $\mathscr{R}^k$. In the end, this yields

$$G_{\mathcal{R}}^{k} = G_{\mathcal{R}}^{k-m} + \left(\mathscr{X}^{k} - G_{\mathcal{R}}^{k-m} \mathscr{R}^{k}\right) \left(\mathscr{R}^{k\mathrm{T}} \mathscr{R}^{k}\right)^{-1} \mathscr{R}^{k\mathrm{T}} \tag{7.71}$$

a rank-m update formula. Note that $\operatorname{rank}\left(\mathscr{R}^{k}\right) = m$. The update formula for $\mathbf{x}^{k+1}$ is

$$\mathbf{x}^{k+1} = \mathbf{x}^{k} - G_{\mathcal{R}}^{k} \mathcal{R}^{k} \tag{7.72}$$

$$= \mathbf{x}^{k} - G_{\mathcal{R}}^{k-m} \mathcal{R}^{k} - \left(\mathscr{X}^{k} - G_{\mathcal{R}}^{k-m} \mathscr{R}^{k}\right) \left(\mathscr{R}^{k\mathrm{T}} \mathscr{R}^{k}\right)^{-1} \mathscr{R}^{k\mathrm{T}} \mathcal{R}^{k} \tag{7.73}$$

$$= \mathbf{x}^{k} - G_{\mathcal{R}}^{k-m} \mathcal{R}^{k} - \left(\mathscr{X}^{k} - G_{\mathcal{R}}^{k-m} \mathscr{R}^{k}\right) \gamma_{k} \tag{7.74}$$

where the column vector γ_k is obtained by solving the normal equations $\left(\mathscr{R}^{k\mathrm{T}} \mathscr{R}^{k}\right) \gamma_k = \mathscr{R}^{k\mathrm{T}} \mathcal{R}^{k}$, which is equivalent to solving the least squares problem

$$\min_{\gamma} \left\| \mathscr{R}^{k} \gamma - \mathcal{R}^{k} \right\|_{2}. \tag{7.75}$$

Note that in Equation (7.74), if $\mathscr{R}^k$ is square and of full rank, then for any $G_{\mathcal{R}}^{k-m}$,

$$\mathbf{x}^{k+1} = \mathbf{x}^{k} - \mathscr{X}^{k} \mathscr{R}^{k^{-1}} \mathcal{R}^{k}, \tag{7.76}$$

the same form as that in the standard secant method.

7.4.2.2 Anderson mixing

The Anderson mixing scheme [5] takes the latest m steps into account to obtain a better approximation to $\mathbf{x}_{n+1}$ without evaluating $\mathcal{R}$ again. Consider

$$\bar{\mathbf{x}}^k = \mathbf{x}^k - \sum_{i=k-m}^{k-1} \gamma_i^k \Delta \mathbf{x}^i = \mathbf{x}_k - \mathscr{X}^k \gamma^k, \tag{7.77}$$

$$\bar{\mathcal{R}}^k = \mathcal{R}^k - \sum_{i=k-m}^{k-1} \gamma_i^k \Delta \mathcal{R}^i = \mathcal{R}^k - \mathscr{R}^k \gamma^k, \tag{7.78}$$

where $\Delta \mathbf{x}^i = \mathbf{x}^{i+1} - \mathbf{x}^i$ and $\Delta \mathcal{R}^i = \mathcal{R}^{i+1} - \mathcal{R}^i$, $\mathscr{X}^k = \left[\Delta \mathbf{x}^{k-m} \cdots \Delta \mathbf{x}^{k-1}\right]$, $\mathscr{R}^k = \left[\Delta \mathcal{R}^{k-m} \cdots \Delta \mathcal{R}^{k-1}\right]$, and $\gamma^k = \left[\gamma_{k-m}^k \cdots \gamma_{k-1}^k\right]^{\mathrm{T}}$. Expressing the equations in the form $\bar{\mathbf{x}}^k = \sum_{j=k-m}^k w_j \mathbf{x}^j$ and $\bar{\mathcal{R}}^k = \sum_{j=k-m}^k w_j \mathcal{R}^j$, it is found that $\sum_{j=k-m}^k w_j = 1$. In other words, $\bar{\mathbf{x}}_k$ and $\bar{\mathcal{R}}_k$ are weighted averages of $\mathbf{x}_{k-m}, \ldots, \mathbf{x}_k$ and $\mathcal{R}^{k-m}, \ldots, \mathcal{R}^k$, respectively. The arguments $\gamma^k = \left[\gamma_{k-m}^k \cdots \gamma_{k-1}^k\right]^{\mathrm{T}}$ are determined by minimizing

$$E\left(\gamma^k\right) = \left\langle \bar{\mathcal{R}}^k, \bar{\mathcal{R}}^k \right\rangle = \left\| \mathcal{R}^k - \mathscr{R}^k \gamma^k \right\|_2^2 \tag{7.79}$$

whose solution can, but should not in practice, be obtained by solving the normal equations

$$\left(\mathscr{R}^{k\mathrm{T}} \mathscr{R}^k\right) \gamma^k = \mathscr{R}^{k\mathrm{T}} \mathcal{R}^k. \tag{7.80}$$

Combining Equations (7.77), (7.78), and (7.80), one obtains

$$\mathbf{x}^{k+1} = \bar{\mathbf{x}}^k + \beta \bar{\mathcal{R}}^k \tag{7.81}$$

$$= \mathbf{x}^k + \beta \mathcal{R}^k - \left(\mathscr{X}^k + \beta \mathscr{R}^k\right) \gamma^k \tag{7.82}$$

$$= \mathbf{x}^k + \beta \mathcal{R}^k - \left(\mathscr{X}^k + \beta \mathscr{R}^k\right) \left(\mathscr{R}^{k\mathrm{T}} \mathscr{R}^k\right)^{-1} \mathscr{R}^{k\mathrm{T}} \mathcal{R}^k \tag{7.83}$$

where β is the preset mixing parameter and $\mathscr{R}^{k\mathrm{T}} \mathscr{R}^k$ is assumed to be nonsingular. In particular, if no previous iterate is taken into account (i.e. $m = 0$), then Equation (7.83) reads

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \beta \mathcal{R}^k \tag{7.84}$$

This scheme is referred to as simple mixing and underrelaxation if $0 < \beta < 1$ (see Section 7.3.3). The update formula (7.83) is the same as (7.74) by setting $G_{\mathcal{R}}^{k-m} = -\beta \mathbf{I}$. In this respect Anderson mixing implicitly forms an approximate inverse Jacobian $G_{\mathcal{R}}^k$ that minimizes $\left\| G_{\mathcal{R}}^k + \beta \mathbf{I} \right\|_F$ subject to (7.68). In the context of mixing, generalized Broyden's second method is equivalent to Anderson mixing. Note that if $\mathscr{R}^k$ is square and nonsingular, then Equation (7.83) matches the formula of the standard secant method.

7.4.2.3 Generalized Broyden's family

Now we can write down the generalized Broyden family, in which an update algorithm is in the form

$$G_{\mathcal{R}}^k = G_{\mathcal{R}}^{k-m} + \left(\mathscr{X}^k - G_{\mathcal{R}}^{k-m} \mathscr{R}^k\right) V^{k\mathrm{T}} \tag{7.85}$$

where $V^{k^{\mathrm{T}}}\mathscr{R}^k = I$ so that the secant condition $G^k_{\mathcal{R}}\mathscr{R}^k = \mathscr{X}^k$ holds. The two optimal choices of $V^{k^T} = M^{k^{-1}}N^{k^T}$ are

$$M^k = \mathscr{R}^{k^T}\mathscr{R}^k, \quad N^{k^T} = \mathscr{R}^{k^T}, \tag{7.86}$$

minimizaing $\left\| G^k_{\mathcal{R}} - G^{k-m}_{\mathcal{R}} \right\|_F$ and

$$M^k = \mathscr{X}^{k^T} G^k_{\mathcal{R}}\mathscr{R}^k, \quad N^{k^T} = \mathscr{X}^{k^T} G^k_{\mathcal{R}}, \tag{7.87}$$

minimizaing $\left\| J^k_{\mathcal{R}} - J^{k-m}_{\mathcal{R}} \right\|_F$. This last choice yields as the approximation for the Jacobian

$$J^k_{\mathcal{R}} = J^{k-m}_{\mathcal{R}} + \left(\mathscr{R}^k - J^{k-m}_{\mathcal{R}}\mathscr{X}^k\right)\left(\mathscr{X}^{k^{\mathrm{T}}}\mathscr{X}^k\right)^{-1}\mathscr{X}^{k^{\mathrm{T}}}, \tag{7.88}$$

after applying the Woodbury formula. The first choice is said to correspond to a Type-II update and the second to a Type-I update (**?**).

7.4.2.4 Anderson's family

The udpate formula for Anderson's family can be found from Equation (7.85) using as the approximation to the previous Jacobian the identity matrix multiplied by a constant β, i.e.,

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \beta\mathcal{R}^k - (\mathscr{X}^k + \beta\mathscr{R}^k)\mathbf{V}^{k^T}\mathcal{R}^k. \tag{7.89}$$

The two choices for $\mathbf{V}^k$ remain the same, replacing $G^{k-m}_{\mathcal{R}}$ by $-\beta\mathbf{I}$. They now minimize $\|G^k_{\mathcal{R}} + \beta\mathbf{I}\|$ and $\|J^k_{\mathcal{R}} + (1/\beta)\mathbf{I}\|$.

7.4.2.5 The Broyden-like class

Both the generalized Broyden's family and Anderson's family can be understood as methods in the Broyden-like class as described in **?**. Ssuppose the latest m iterates are available, which are denoted by $\mathbf{x}^1,\ldots,\mathbf{x}^m$. Let $\Delta\mathbf{x}^i = \mathbf{x}^{i+1} - \mathbf{x}^i$ for $i = 1,\ldots,m-1$. Partition $\Delta\mathbf{x}^1,\ldots,\Delta\mathbf{x}^{m-1}$ into p groups,

$$\mathscr{X}^1 = \left[\Delta\mathbf{x}^1,\ldots,\Delta\mathbf{x}^{z_1}\right], \tag{7.90}$$

$$\mathscr{X}^2 = \left[\Delta\mathbf{x}^{z_1+1},\ldots,\Delta\mathbf{x}^{z_2}\right], \tag{7.91}$$

$$\vdots \tag{7.92}$$

$$\mathscr{X}^p = \left[\Delta\mathbf{x}^{z_{k-1}+1},\ldots,\Delta\mathbf{x}^{z_k}\right], \tag{7.93}$$

where z_i is the index of the last entry in the i th group for $i = 1,\ldots,p$; $z_0 = 0$ and $z_p = m-1$. Also partition $\Delta\mathcal{R}^1,\ldots,\Delta\mathcal{R}^{m-1}$ into $\mathscr{R}^1,\ldots,\mathscr{R}^p$ accordingly, where $\Delta\mathcal{R}^i = \mathcal{R}^{i+1} - \mathcal{R}^i$ with $\mathcal{R}^i = \mathcal{R}(\mathbf{x}^i)$. The sizes of the groups for $i = 1,\ldots,k$ are denote by $s_i := z_i - z_{i-1}$. Note that the indexing here is different from the previous sections. The inverse of the Jacobian is iteratively approximated at the (z_i+1)st iterate for $i = 1,\ldots,p$ as

$$G^{i+1}_{\mathcal{R}} = G^i_{\mathcal{R}} + \left(\mathscr{X}^i - G^i_{\mathcal{R}}\mathscr{R}^i\right)\mathbf{V}^{i^T}, \tag{7.94}$$

where $\mathbf{V}^{i\,T}\mathscr{R}^i = \mathbf{I}$ for the secant condition. The update follows the formula of the generalized Broyden family. In the context of mixing, the base case is

$$G^1_{\mathcal{R}} = -\beta\mathbf{I}, \tag{7.95}$$

where β is the mixing parameter. The next iterate is set as

$$\mathbf{x}^{m+1} = \mathbf{x}^m - G^{k+1}_{\mathcal{R}}\mathcal{R}^m. \tag{7.96}$$

The choice of V_i satisfying $V_i^{\mathrm{T}}\mathscr{F}_i = I$ is performed as described in Section 7.4.2.3.

7.4.3 Practical considerations

The application of Broyden's method as described so far is unfeasible for the problems considered in this document, i.e., thermo-mechanical coupled problems with many unknowns. So far, the descriptions considered of the Broyden-like class methods require one to keep the large $G^i_{\mathcal{R}}$ matrices of size $n_{\text{unknowns}} \times n_{\text{unknowns}}$ in memory, in addition to the previous iterates. This is a significant drawback. However, **?** present a more memory efficient way of implementing these methods. Let

$$E^i = \mathscr{X}^i - G^i_{\mathcal{R}}\mathscr{R}^i. \tag{7.97}$$

Substituting Equation (7.97) into Equation (7.94) one obtains

$$\begin{aligned} G^i_{\mathcal{R}} &= G^{i-1}_{\mathcal{R}} + E^{i-1}V^{i-1\,T}, && (7.98)\\ &= G^{i-2}_{\mathcal{R}} + E^{i-2}V^{i-2\,T} + E^{i-1}V^{i-1\,T}, && (7.99)\\ &\vdots && (7.100)\\ &= G^1_{\mathcal{R}} + \sum_{j=1}^{i-1} E^j V^{j\,\mathrm{T}}, && (7.101) \end{aligned}$$

for $i = 2,\ldots,p+1$. Matrices $G^i_{\mathcal{R}}$ need not be explicitly stored. $G^i_{\mathcal{R}}$ is needed only to compute $G^i_{\mathcal{R}}\mathscr{R}^i$ in Equation (7.97) and $G^{p+1}_{\mathcal{R}}\mathcal{R}^m$ in Equation (7.96), and also for V^i if it depends on $G^i_{\mathcal{R}}$. Substituting Equation (7.101) into Equation (7.97) obtains

$$E^i = \mathscr{X}^i - G^1_{\mathcal{R}}\mathscr{R}^i - \sum_{j=1}^{i-1} E^j\left(V^{j\,T}\mathscr{R}^i\right), \tag{7.102}$$

for $i = 1,\ldots,p$. The computation is economic for large-scale problems with $n \gg m$. The next iterate $\mathbf{x}^{m+1}$ in (32) can also be computed in a similar manner

$$\mathbf{x}^{m+1} = \mathbf{x}^m - G^{p+1}_{\mathcal{R}}\mathcal{R}^m = \mathbf{x}^m - G^1_{\mathcal{R}}\mathcal{R}^m - \sum_{j=1}^{k} E^j\left(V^{j\,T}\mathcal{R}^m\right). \tag{7.103}$$

Using Type-II update, the computation of V^i is straightforward from $\mathscr{R}^i$. On the other hand, Type-I update involves $G^i_{\mathcal{R}}$ to compute V^i. Thus

$$N^{i\,T} = \mathscr{X}^{i\,T}G^i_{\mathcal{R}} = G^1_{\mathcal{R}}\mathscr{X}^{i\,T} + \sum_{j=1}^{i-1}\left(\mathscr{X}^{i\,\mathrm{T}}E^j\right)V^{j\,T}. \tag{7.104}$$

After obtaining N^i, we compute $M^i = {N^i}^T \mathscr{R}^i$ and then ${V^i}^T = {M^i}^{-1} {N^i}^T$ for $i = 1, \ldots, p$.

Looking at Equations (7.102), (7.103) and (7.104), one still needs the initial approximation to the inverse of the Jacobian, $G^1_{\mathcal{R}}$, whose size is $n_{\text{unknown}} \times n_{\text{unknown}}$. In **?** the approach adopted was to follow the idea of Anderson's mixing and set

$$G^1_{\mathcal{R}} = -\beta \mathbf{I}, \tag{7.105}$$

drastically improving the memory requirements, as only one scalar parameter, β, needs to be saved. Also, **?**, assumes in his implementation of Broyden's method, an initial approximation to $G^1_{\mathcal{R}}$ equal to the identity matrix. Information about $G^1_{\mathcal{R}}$ is applied in the preconditioning of the system instead.

To compute V^i, **?** suggest a QR decomposition with pivoting. Be it for a Type-II update, where one needs to solve a least-squares problem, or for a Type-I update, one needs to invert a generally non-symmetric matrix. This approach leads to better numerical stability when the matrices to be inverted are singular or ill-conditioned, compared with solving the normal equations. The QR decomposition has a computational cost of $\mathcal{O}(n_{\text{unknown}} m^2)$ algebraic operations (**?**).

If the size of the groups $s_1, s_2, \ldots$ are fixed from one Newton iteration to the next, so the E^i and V^i matrices remain the same from one iteration to the next, the computation effort to compute them is saved from one iteration to the next. Here only constant $s = s_1 = s_2 = \cdots$ is considered, where $s = \infty$ corresponds to Anderson's mixing, where all previous iterates available are considered.

One question remains. How to proceed when the available memory runs out?. According to **?**, as is often the case with GMRES, the iteration can be restarted if there is no more room to store the vectors. A different approach, called limited memory in the optimization literature, is to replace the oldest of the stored steps with the most recent one.

Thus, appplying the method as suggest by **?**, leads to a storage need of

1. Two column vectors of size n_{unknown} for $\mathbf{x}^m$ and $\mathcal{R}^m$.
2. An $n_{\text{unknown}} \times (m-1)$ matrix for $\mathscr{X}^1, \ldots, \mathscr{X}^k$ (shared with $E^1, \ldots, E^k$).
3. An $n_{\text{unknown}} \times (m-1)$ matrix for $\mathscr{R}^1, \ldots, \mathscr{R}^k$ (shared with $V^1, \ldots, V^k$).
4. For Type-I update we also store the last group N^k, since its computation involves $G^k_{\mathcal{R}}$.

and for each nonlinear iteration the computational cost is $\mathcal{O}(n_{\text{unknown}} m^2)$. For $m = 1$, V^i can be directly computed without needing to invert matrices, and the cost comes down to $\mathcal{O}(n_{\text{unknown}})$ per nonlinear iteration. Also, when k is different from any z_i a group is being complete, one can either use a shortened group or reuse the approximation to the Jacobian without using the new information. This save computational effort and those iteration cost only $\mathcal{O}(n_{\text{unknown}})$. See in Box 7.5 the pseudocode for the Broyden-like family with restart.

? presents for the Broyden method an implementation that halves the memory requirement relative to the one present in **?**. It is based on the Type-I update and to deduce it consider Equation (7.61) and Sherman-Morrison formula

$$\left(J_{\mathcal{R}} + \mathbf{u}\mathbf{v}^T\right)^{-1} = \left(\mathbf{I} - \frac{\left(G_{\mathcal{R}}\mathbf{u}\right)\mathbf{v}^T}{1 + \mathbf{v}^T G_{\mathcal{R}} \mathbf{u}}\right) G_{\mathcal{R}}, \tag{7.106}$$

where as before $G_{\mathcal{R}} \equiv J_{\mathcal{R}}^{-1}$. One can rewrite Equation (7.61) as

$$J_{\mathcal{R}}^{k+1} = J_{\mathcal{R}}^{k} + \mathbf{u}^k \mathbf{v}^{k\,T}, \tag{7.107}$$

where

$$\mathbf{u}^k = \left(\Delta\mathcal{R}^k - J_{\mathcal{R}}^k \Delta\mathbf{x}^k\right) / \left\|\Delta\mathbf{x}^k\right\| \text{ and } \mathbf{v}^k = \Delta\mathbf{x}^k / \left\|\Delta\mathbf{x}^k\right\|. \tag{7.108}$$

Then, keeping in mind that $J_{\mathcal{R}}^1 = \mathbf{I}$,

$$G_{\mathcal{R}}^{k+1} = \left(\mathbf{I} - \mathbf{w}^k \mathbf{v}^{k\,T}\right)\left(\mathbf{I} - \mathbf{w}_{k-1} \mathbf{v}^{k-1\,T}\right) \cdots \left(\mathbf{I} - \mathbf{w}^1 \mathbf{v}^{1\,T}\right) G_{\mathcal{R}}^1, \tag{7.109}$$

$$= \prod_{j=0}^{k} \left(\mathbf{I} - \mathbf{w}^j \mathbf{v}^{j\,T}\right), \tag{7.110}$$

where, for $k \geq 0$,

$$\mathbf{w}^k = \frac{G_{\mathcal{R}}^k \mathbf{u}^k}{1 + \mathbf{v}^{k\,T} G_{\mathcal{R}}^k \mathbf{u}^k}. \tag{7.111}$$

So, to apply $G_{\mathcal{R}}^{k+1}$ to a vector $\mathbf{p}$, the is cost of $O(n_{\text{unknown}} k)$ floating point operations and storage of the $2k$ vectors $\left\{\mathbf{w}^j\right\}_{j=1}^{k}$ and $\left\{\Delta\mathbf{x}^j\right\}_{j=1}^{k}$. The storage can be halved with a trick (see **?** for details)

$$\Delta\mathbf{x}^k = -G_{\mathcal{R}}^{k+1} \mathcal{R}^k, \tag{7.112}$$

$$= -\left(\mathbf{I} - \frac{\mathbf{w}^k \Delta\mathbf{x}^{k\,T}}{\left\|\Delta\mathbf{x}^k\right\|}\right) G_{\mathcal{R}}^k \mathcal{R}^k, \tag{7.113}$$

$$= -\frac{G_{\mathcal{R}}^k R^k}{1 + \Delta\mathbf{x}^{k\,T} G_{\mathcal{R}}^k \mathcal{R}^k / \|\Delta\mathbf{x}^k\|^2}. \tag{7.114}$$

According to **?**, the Sherman-Morrison approach is more efficient, in terms of both time and storage, than dense matrix approaches proposed elsewhere. For example, the approach presented in **?** has a $\mathcal{O}(n_{\text{unknown}})$ cost per nonlinear iteration and requires one to keep in memory the QR decomposition of the previous approximation to the Jacobian. However, the dense matrix approach can detect ill-conditioning in the approximate Jacobians. Bounded deterioration implies that the Broyden matrices will be well-conditioned if the data is sufficiently good, and superlinear convergence suggests that only a few iterates will be needed.

In the context of FSI The multi-secant quasi-Newton methods have been used in the context of FSI, although not always presented as such (**????**). **?** and **?** consider the system of equations (7.45) and (7.46), where recall that an estimate for the Jacobians $J_{\mathcal{U}}$ and $J_{\mathcal{T}}$ is needed. The authors achieve this by using linear reduced-order models for the fluid solver and the structure solver. These are set up from solver input and output deltas or sensitivities during the coupling iterations. The resulting method for two black-box solvers is called interface block quasi-Newton method with least-squares approximation (IBQN-LS) in **?**.

This approach can be understood in the framework of the multi-secant quasi-Newton methods presented above and originating in **?** as follows. If one looks at

$\beta\mathbf{I} - (\tilde{\mathcal{X}}^k + \beta\tilde{\mathcal{R}}^k)\mathbf{V}^{k^T}$ in Equation (7.89) as, in a sense, an approximation to the inverse of the Jacobian (compare with Equation (7.85)). The corresponding Jacobian is given by

$$J_{\mathcal{R}}^k = \alpha\mathbf{I} + (\tilde{\mathcal{R}}^k - \alpha\mathbf{I}\tilde{\mathcal{X}}^k)(\tilde{\mathcal{X}}^{k^T}\tilde{\mathcal{X}}^k)^{-1}\tilde{\mathcal{X}}^{k^T}, \tag{7.115}$$

where $\alpha = 1/\beta$. If one sets $\alpha = 0$, the approximation to the Jacobian obtained is

$$J_{\mathcal{R}}^k = \tilde{\mathcal{R}}^k(\tilde{\mathcal{X}}^{k^T}\tilde{\mathcal{X}}^k)^{-1}\tilde{\mathcal{X}}^{k^T}. \tag{7.116}$$

This corresponds to the linear reduced order models in **?**, where $\mathcal{R}$ is replaced by the functions corresponding to the fluid and structure solvers. If the functions considered are instead the mechanical and thermal solvers, this method can easily be applied to the thermomechanical problem. The block $(\tilde{\mathcal{X}}^{k^T}\tilde{\mathcal{X}}^k)^{-1}\tilde{\mathcal{X}}^{k^T}$ can be understood as being a part of a least-squares solution, the so-called normal equations, i.e., the equations whose solution also solve the minimization problem

$$\arg\min_{\tilde{\gamma}} \|\Delta\mathbf{x} - \tilde{\mathcal{X}}^k\tilde{\gamma}\|_2. \tag{7.117}$$

As such one can avoid the use of the normal equations and employ more numerically stable and efficient methods such economy size QR-decomposition. In addition, **?** solve the system of equation (7.45) and (7.46) in a Gauss-Seidel manner, using always the most recent values available to estimate the Jacobians.

An interface quasi-Newton method based on Equation (7.38) and Equation (7.39) is presented in **?** for FSI. The method is called interface quasi-Newton with an approximation of the inverse of the interface Jacobian matrix by least squares (QIN-ILS). Its origin is the IBQN-LS method presented in **?**, and it employs only one reduced-order model for the inverse of the overall interface Jacobian matrix of the Newton system (Equation (7.38)) applied to the right-hand side vector.

If in Equation (7.83), corresponding to Anderson's mixing, one sets $\beta = -1$, the update formula comes out to be

$$\mathbf{x}^{k+1} = \mathbf{x}^k - \mathcal{R}^k - (\tilde{\mathcal{X}}^k - \tilde{\mathcal{R}}^k)\left(\tilde{\mathcal{R}}^{k^{\mathrm{T}}}\tilde{\mathcal{R}}^k\right)^{-1}\tilde{\mathcal{R}}^{k^{\mathrm{T}}}\mathcal{R}^k. \tag{7.118}$$

Using the definition for the fixed-point function $\mathcal{S}$, one finds

$$\mathbf{x}^{k+1} = \mathcal{S}^k - \mathcal{S}^k\left(\tilde{\mathcal{R}}^{k^{\mathrm{T}}}\tilde{\mathcal{R}}^k\right)^{-1}\tilde{\mathcal{R}}^{k^{\mathrm{T}}}\mathcal{R}^k, \tag{7.119}$$

or

$$\Delta\mathbf{x}^k = -\mathcal{R}^k - \mathcal{S}^k\left(\tilde{\mathcal{R}}^{k^{\mathrm{T}}}\tilde{\mathcal{R}}^k\right)^{-1}\tilde{\mathcal{R}}^{k^{\mathrm{T}}}\mathcal{R}^k, \tag{7.120}$$

When no delta columns are available yet, constant relaxation is used once to ensure stability.

Box 7.5: Broyden-like class methods for one timestep with restart.

(i) Set $\mathbf{x}^1 = \mathbf{x}^p_{n+1}$.

(ii) Evaluate $\mathcal{R}^1 = \mathcal{R}(\mathbf{x}^1)$, which implies the solution of the mechanical and the thermal problems, $\mathcal{U}$ and $\mathcal{T}$, respectively.

(iii) Compute $\mathbf{x}^2 = \mathbf{x}^1 + \beta R^1$.

(iv) Evaluate $\mathcal{R}^2 = \mathcal{R}(\mathbf{x}^2)$, which implies the solution of the mechanical and the thermal problems, $\mathcal{U}$ and $\mathcal{T}$, respectively.

(v) Intialize counters: $k = 2$ and $i = 1$

(vi) Enter the nonlinear loop

(1) If the maximum number of previous iteration m has been reached restart all $\mathscr{X}^i$ and $\mathscr{R}^i$ and set $i = 1$.

(2) Compute $\Delta\mathcal{R}^{k-1} = \mathcal{R}^k - \mathcal{R}^{k-1}$ and add it to $\mathscr{R}^i$.

(3) Compute $\Delta\mathbf{x}^{k-1} = \mathbf{x}^k - \mathbf{x}^{k-1}$ and add it to $\mathscr{X}^i$.

(4) If $s = \infty$ (Anderson mixng):

- Compute ${V^i}^T$ according to the type of update chosen (Equation (7.87) and (7.86)).
- Update according to $\mathbf{x}^{k+1} = \mathbf{x}^k + \beta\mathcal{R}^k - (\mathscr{X}^i + \beta\mathscr{R}^i){V^i}^T\mathcal{R}^k$.

(5) else:

- If $k = z_j + 1$ for any $j \geq 1$, compute E^i (Equation (7.102)) and ${V^i}^T$ according to the type of update chosen (Equation (7.87) and (7.86)). Save them on $\mathscr{X}^i$ and $\mathscr{R}^i$, respectively. Update $i = i + 1$.
- Update according to $\mathbf{x}^{k+1} = \mathbf{x}^k + \beta\mathcal{R}^k - \sum_{j=1}^{i-1} E^j({V^j}^T\mathcal{R}^k)$.

(6) Evaluate $\mathcal{R}^{k+1} = \mathcal{R}(\mathbf{x}^{k+1})$, which implies the solution of the mechanical and the thermal problems, $\mathcal{U}$ and $\mathcal{T}$, respectively.

(7) If the desired accuracy has not been reached, update $k = k+1$ and go to step (1).

7.5 Multipoint iteration functions

7.5.1 Finite-Difference Newton Method

This approach follows precisely the one described in Section 7.3.2 for the "standard" Newton method. The difference lies in the computation of the Jacobian. Here the Jacobian $J_{\mathcal{R}}(\mathbf{x})$ is approximated from a forward finite-difference, $J^h_{\mathcal{R}}(\mathbf{x})$, by columns. Following **?**. the jth column is

$$\left[J^h_{\mathcal{R}}(\mathbf{x})\right]_j = \begin{cases} \dfrac{\mathcal{R}(\mathbf{x}+h\sigma_j\mathbf{e}_j)-\mathcal{R}(\mathbf{x})}{\sigma_j h}, & x_j \neq 0 \\ \dfrac{\mathcal{R}(h\mathbf{e}_j)-\mathcal{R}(\mathbf{x})}{h}, & x_j = 0 \end{cases}, \tag{7.121}$$

where $\mathbf{e}_j$ is the unit vector in the j th coordinate direction. The difference increment h should be no smaller than the square root of the inaccuracy in $\mathcal{R}$ (**?**). It should, however, be scaled. Rather than simply perturbing $\mathbf{x}$ by a difference increment h, roughly the square root of the error in $\mathcal{R}$, in each coordinate direction, the perturbation is multiplied to compute the j th column by

$$\sigma_j = \max(|(x)_j|, 1)\,\mathrm{sign}((x)_j), \tag{7.122}$$

with a view toward varying the correct fraction of the low-order bits in $(x)_j$. The sign function is defined as

$$\mathrm{sgn}(z) = \begin{cases} z/|z| & \text{if } z \neq 0 \\ 1 & \text{if } z = 0 \end{cases}. \tag{7.123}$$

While this scaling usually makes little difference, it can be crucial if $|(x)_j|$ is very large. Note that there is no adjustment if $|(x)_j|$ is very small because the error determines the lower limit on the size of the difference increment in $\mathcal{R}$. For example, if evaluations of $\mathcal{R}$ are accurate to 16 decimal digits, the difference increment should change roughly the last eight digits of x. (**?**)

Each column of $J^h_{\mathcal{R}}(\mathbf{x})$ requires one new function evaluation and, therefore, a finite difference Jacobian costs n_{unknown} function evaluations. If the perturbation is appropriately chosen, the method converges quadratically when the function satisfies certain conditions, and the initial attempt is close enough to the solution (**?**).

The Chord and Shamanskii Methods If the computational cost of a forward difference Jacobian is high, i.e., $\mathcal{R}$ is expensive and/, or n_{unknown} is significant. If an analytic Jacobian is not available, it is wise to amortize this cost over several nonlinear iterations. The chord method does precisely that. It differs from Newton's method in that the evaluation and factorization of the Jacobian are done only once for $J_{\mathcal{R}}(\mathbf{x}^0)$. The advantages of the chord method increase as n increases, since both the n function evaluations and the $O(n^3_{\text{unknown}})$ work (in the dense matrix case) in the matrix factorization are done only once. So, while the convergence is q-linear and more nonlinear iterations will be needed than for Newton's method, the overall cost of the solution will usually be much less. A middle ground is the Shamanskii method. Here the Jacobian factorization and matrix function evaluation is done after every m computations of the step (**?**).

Since the present use-case, the number of unknowns n is very large, and the evaluation of the function $\mathcal{R}$ is also costly, approximating the Jacobian using a finite difference is not suitable, even utilizing the chord or Shamanskii methods.

7.5.2 Newton-Krylov methods

In the Newton-Krylov methods, the solution of the Newton system of equations in Equation (7.38) is achieved using Krylov methods, such as GMRES or BiCGSTAB. The Krylov iterative methods approximate the solution of a linear system $\mathbf{Ax} = \mathbf{b}$ using the Krylov subspace

$$\mathcal{K}_m = \text{span}\{\mathbf{r}_0, \mathbf{A}\mathbf{r}_0, \mathbf{A}^2\mathbf{r}_0, \ldots, \mathbf{A}^{m-1}\mathbf{r}_0\}, \tag{7.124}$$

such that the mth iterate, $\mathbf{x}_m \in \mathcal{K}_m$, with $\mathbf{r}_0 = \mathbf{b} - \mathbf{A}\mathbf{x}_0$. The precise way the $\mathbf{x}_m$ is built is what distinguishes the different methods.

To produce the appropriate Krylov subspace, one needs the product $J_{\mathcal{R}}(\mathbf{x}^k)\mathbf{y}$ in Equation (7.38), for some vector $\mathbf{y}$. It is assumed that the Jacobian is not available, so it must be approximated. Also, it would be beneficial if the full Jacobian is neither computed in its entirety nor wholly stored in memory, i.e., a matrix-free method is desirable. As in Section 7.5.1, the Jacobian-vector product is easy to approximate with a forward difference directional derivative (**?**). The forward difference directional derivative at $\mathbf{x}^k$ in the direction $\mathbf{q}$ is

$$J^h_{\mathcal{R}}(\mathbf{x}^k)\mathbf{q} = \begin{cases} \mathbf{0}, & \mathbf{q} = \mathbf{0} \\ \|\mathbf{q}\| \dfrac{\mathcal{R}(\mathbf{x}^k + \sigma(\mathbf{x}^k, \mathbf{q}) h \mathbf{q}/\|\mathbf{q}\|) - \mathcal{R}(\mathbf{x}^k)}{\sigma(\mathbf{x}^k, \mathbf{q}) h}, & \mathbf{q} \neq \mathbf{0}. \end{cases} \tag{7.125}$$

The scaling is important. $\mathbf{q}$ is scaled to be a unit vector and take a numerical directional derivative in the direction $\mathbf{q}/\|\mathbf{q}\|$. If h is roughly the square root of the error in $\mathcal{R}$, a difference increment in the forward difference is used to make sure that the appropriate low-order bits of $\mathbf{x}^k$ is perturbed. So h is multiplied by

$$\sigma(\mathbf{x}^k, \mathbf{q}) = \max(|{\mathbf{x}^k}^T \mathbf{q}|, \|\mathbf{q}\|) \, \text{sign}({\mathbf{x}^k}^T \mathbf{q}) / \|\mathbf{q}\|. \tag{7.126}$$

? describes two different Newton-Krylov methods, the Newton-Arnoldi and the Newton-GMRES. The goal is to solve Equation (7.38) and find $\Delta\mathbf{x}^k$. The Krylov iterates are denoted by $\Delta\mathbf{x}^*_m$.

The first phase of both algorithms is identical. They both use the Arnoldi-Gram-Schmidt process to produce an orthogonal basis for the Krylov subspace, $\{\mathbf{q}_1, \ldots, \mathbf{q}_m\}$ for some integer k. Consider the unitary matrix $\mathbf{Q}_m$

$$\mathbf{Q}_m = [\mathbf{q}_1 \cdots \mathbf{q}_m], \tag{7.127}$$

and the upper Hessenber matrix $\mathbf{H}_m$ and related matrix $\bar{\mathbf{H}}_m$

$$\mathbf{H}_m = \begin{bmatrix} h_{11} & h_{12} & \cdots & \cdots & h_{1m} \\ h_{21} & h_{22} & \cdots & \cdots & h_{2m} \\ & \ddots & \ddots & & \vdots \\ & & \ddots & \ddots & \vdots \\ & & & h_{m,m-1} & h_{mm} \end{bmatrix}, \quad \bar{\mathbf{H}}_m = \begin{bmatrix} h_{11} & h_{12} & \cdots & \cdots & h_{1m} \\ h_{21} & h_{22} & \cdots & \cdots & h_{2m} \\ & \ddots & \ddots & & \vdots \\ & & \ddots & \ddots & \vdots \\ & & & \ddots & h_{mm} \\ & & & & b_{m+1,m} \end{bmatrix}, \tag{7.128}$$

where the quantities h_{ij} are also obtained from the algorithm. $\mathbf{H}_m$ can be interpreted as the projection of $\mathbf{A}$ onto the Krylov subspace since $\mathbf{H}_m = \mathbf{Q}_m^T J_{\mathcal{R}}(\mathbf{x}^k)\mathbf{Q}_m$. See Box 7.6 for the pseudo-code of the Arnoldi-Gram-Schmidt process.

Box 7.6: Arnoldi process to orthonormalize the Krylov subspace

(i) Compute $\mathbf{r}_0 = -\mathcal{R}(\mathbf{x}^k)$ and set $\beta = \|\mathbf{r}_0\|$ and $\mathbf{q}_1 = \mathbf{r}_0/\beta$.

(ii) $j = 1$

(iii) $\mathbf{a}_{j+1}^{(1)} = J_{\mathcal{R}}(\mathbf{x}^k)\mathbf{q}_j$.

(iv) Compute $h_{ij} = (\mathbf{q}_i, \mathbf{a}_{j+1}^{(i)})$ and compute $\mathbf{a}_{j+1}^{(i+1)} = \mathbf{a}_{j+1}^{(i)} - h_{ij}\mathbf{q}_i$ for $i = 1, \ldots, j$.

(v) Compute $h_{j+1,j} = \|\mathbf{a}^{(j+1)}{}_{j+1}\|$ and set $\mathbf{q}_{j+1} = \mathbf{a}_{j+1}^{(j+1)}/h_{j+1,j}$.

(vi) $j = j + 1$

(vii) If $j < m - 1$ go to Step (iii)

After obtaining an orthogonal basis for the Krylov subspace $\mathcal{K}_m$, the Newton-Arnoldi method projects the mth residual onto the Krylov subspace and sets it to zero, similar to a weigthed residual method, i.e.,

$$\mathbf{Q}_m \left(\mathcal{R}(\mathbf{x}^k) - J_{\mathcal{R}}(\mathbf{x}^k)\Delta\mathbf{x}_m^* \right) = \mathbf{0}. \tag{7.129}$$

In practice, one solves the equivalent linear system $\mathbf{H}_m\boldsymbol{\eta} = \beta\mathbf{e}_1$ for $\boldsymbol{\eta}$ and set $\Delta\mathbf{x}^k = \Delta\mathbf{x}_0^* + \mathbf{Q}_m\boldsymbol{\eta}$.

The Newton-GMRES attempts to minimize the norm of the mth residual, i.e.,

$$\Delta\mathbf{x}_m^* = \arg\min_{\mathbf{y}\in\mathcal{K}_m} \left\| \mathcal{R}(\mathbf{x}^k) - J_{\mathcal{R}}(\mathbf{x}^k)\mathbf{y} \right\|_2. \tag{7.130}$$

This is exectued in practice solving the linear least-squares problem $\|\beta\mathbf{e}_1 - \bar{\mathbf{H}}_m\boldsymbol{\eta}\|$ for $\boldsymbol{\eta}$ and set $\Delta\mathbf{x}^k = \Delta\mathbf{x}_0^* + \mathbf{Q}_m\boldsymbol{\eta}$.

Since the present use case includes many unknowns, it leads to memory concerns if the Krylov subspace is allowed to grow indefinitely. A restarted version where the maximum size of the Krylov space is restricted to m elements is preferred. Once this number is reached, the procedure is restarted. However, if m is small, the convergence can be poor.

In each nonlinear iteration of the Newton-Krylov, the number of iterations can be large, and each iteration requires an evaluation of the function. This can be a significant drawback when the intended use assumes that evaluating the function $\mathcal{R}$ is expensive. This problem is however mitigated by the fact the Newton system is only solved until it satisfies

$$\|J_{\mathcal{R}}(\mathbf{x}^k)\Delta x_m^* + \mathcal{R}(\mathbf{x}^k)\| \leq \eta\|\mathcal{R}(\mathbf{x}^k)\|, \tag{7.131}$$

where η is called the forcing term and it is chosen to avoid oversolving the Newton system (Equation (7.38). As a simple approach, **?** suggests $\eta = 0.1$. However, he describes more sophisticated ways to choose this parameter, such as the Eisenstat-Walker method. The smaller the forcing term η, the closer one gets to the "standard" Newton method. However, especially in the first nonlinear iterations,

choosing a η that is too small leads to unnecessarily long computational times. The linear system is being solved with too much precision. See Box 7.7 for the pseudocode of the Newton-Arnoldi and Newton-GMRES.

Box 7.7: Timestep n of the Newton-Krylov methods with restart, Newton-Arnoldi and Newton-GMRES.

(i) $k = 0$

(ii) Enter the Newton loop

(1) Compute $\mathbf{r}_0 = -\mathcal{R}(\mathbf{x}^k)$

(2) Enter the Krylov loop

(a) Set $\beta = \|\mathbf{r}_0\|$ and $\mathbf{q}_1 = \mathbf{r}_0/\beta$.

(b) $j = 1$

(c) $\mathbf{a}_{j+1}^{(1)} = J_{\mathcal{R}}(\mathbf{x}^k)\mathbf{q}_j$.

(d) Compute $h_{ij} = (\mathbf{q}_i, \mathbf{a}_{j+1}^{(i)})$ and compute $\mathbf{a}_{j+1}^{(i+1)} = \mathbf{a}_{j+1}^{(i)} - h_{ij}\mathbf{q}_i$ for $i = 1, \ldots, j$.

(e) Compute $h_{j+1,j} = \|\mathbf{a}^{(j+1)}{}_{j+1}\|$ and set $\mathbf{q}_{j+1} = \mathbf{a}_{j+1}^{(j+1)}/h_{j+1,j}$.

(f) If using the Arnoldi method:

- Solve the linear system $\mathbf{H}_j\boldsymbol{\eta} = \beta\mathbf{e}_1$ for $\boldsymbol{\eta}$
- set $\Delta\mathbf{x}_j^* = \Delta\mathbf{x}_0^* + \mathbf{Q}_j\boldsymbol{\eta}$.

(g) Else if using the GMRES method:

- Solve the linear least-squares problem $\|\beta\mathbf{e}_1 - \bar{\mathbf{H}}_j\boldsymbol{\eta}\|$ for $\boldsymbol{\eta}$.
- Set $\Delta\mathbf{x}_j^* = \Delta\mathbf{x}_0^* + \mathbf{Q}_j\boldsymbol{\eta}$.

(h) If $\|J_{\mathcal{R}}(\mathbf{x}^k)\Delta x_j^* + \mathcal{R}(\mathbf{x}^k)\| \leq \eta\|\mathcal{R}(\mathbf{x}^k)\|$ is not satisfied set $j = j + 1$.

(i) if $j > m$, restart the method going to Step (a). Else, go to Step (c).

(3) $\mathbf{x}^{k+1} = \mathbf{x}^k + \Delta x^k$.

(4) $k = k + 1$

(5) $\mathbf{r}_0 = -\mathcal{R}(\mathbf{x}^k)$

(6) If convergence has not been reached, $\|\mathbf{r}_0\| > \epsilon_1$, go to Step (2)

7.5.3 Extrapolation tecniques in cycling mode

There is a vast literature on sequence acceleration/extrapolation methods (see **?** and **?** for textbook treatments of this topic). One often deals with sequences and series in numerical analysis, applied mathematics, and engineering. They are produced by iterative methods, perturbation techniques, and approximation procedures depending on a parameter. Those sequences or series often converge so slowly that it is a severe drawback to their practical use. Convergence acceleration methods present a solution and have been studied for many years and applied to various situations. They are based on the very natural idea of extrapolation. In many cases, they lead to the solution of

unsolvable problems otherwise. Sequences of vectors can also be considered, with their dimension being substantial. Such sequences arise, for example, in the solution by fixed-point iterative methods of systems of linear or nonlinear algebraic equations.

An example of a scalar acceleration method is first presented to fix ideas. Let (S_n) be a sequence of numbers that converges to S. This sequence can be transformed into another, denoted (T_n). For example, consider

$$T_n = \frac{S_n S_{n+2} - S_{n+1}^2}{S_{n+2} - 2S_{n+1} + S_n}, \quad n = 0, 1, \ldots, \tag{7.132}$$

which corresponds to the Aitken Δ^2 process.

This expression can be obtained considering a transformation that would yield the limit of a geometric sequence from only three iterates, i.e., if one fits an exponential function

$$S + a\lambda^n, \tag{7.133}$$

the sequence transformation takes the horizontal asymptote of the exponential, S. For the geometrical interpretation of Aitken's Δ^2 method, see Figure 7.4.

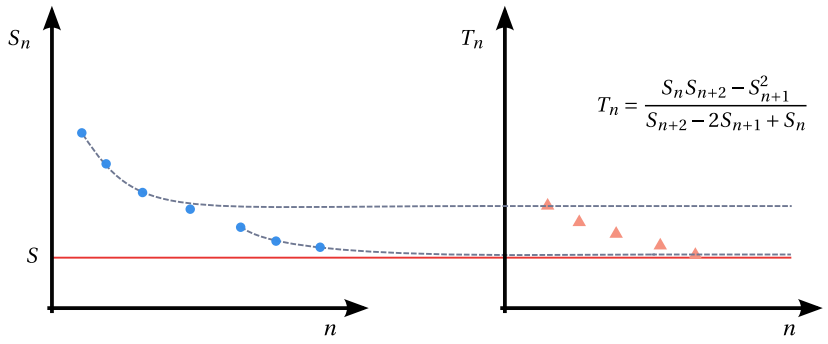


Figure 7.4: Geometrical interpretation of Aitken's Δ^2 method.

One can also show that if (S_n) goes to its limit S at a rate strictly greater than 1[1], (T_n) does not have a better rate of convergence.

In practice, the sequence produced by Aitken's Δ^2 method tends to converge faster to the limit than (S_n) does. Very often, it is much cheaper to calculate (T_n), which involves only the calculation of differences, one multiplication, and one division, than to calculate many more terms of the sequence (S_n). Care must be taken, however, to avoid introducing errors due to insufficient precision when calculating the differences in the numerator and denominator of the expression.

There is, however, no universal sequence accelerator capable of accelerating all sequences. It is also the case that nonlinear transformations can even fail to converge or converge to a value other than the limit of the original sequence (**?**).

According to **?**, there is a very strong connection between sequence transformations and fixed point methods for solving $x = g(x)$, $g\colon \mathbb{R} \to \mathbb{R}$. The most well-known example of this connection is that between Aitken's Δ^2 process and Steffensen's method.

$$T_n = S_n - \frac{(S_{n+1} - S_n)^2}{S_{n+2} - 2S_{n+1} + S_n}, \quad n = 0, 1, \ldots \quad \text{for Aitken's process} \tag{7.134}$$

[1] (S_n), $n \in \mathbb{N}$ converges linearly to S if there exists a number $\mu \in (0,1)$ such that $\lim_{n\to\infty} \frac{|S_{n+1} - S|}{|S_n - S|} = \mu$.

and

$$x_{n+1} = x_n - \frac{\left(g\left(x_n\right) - x_n\right)^2}{g\left(g\left(x_n\right)\right) - 2g\left(x_n\right) + x_n}, \quad n = 0, 1, \ldots \quad \text{for Steffensen's method.} \tag{7.135}$$

Turning to vector sequences and systems of nonlinear equations, let $F : (\mathbf{w}^k) \to (\mathbf{y}^k)$ be a vector extrapolation method defined by

$$\mathbf{y}^k = F\left(\mathbf{w}^k, \ldots, \mathbf{w}^{k+m}\right), \quad n = 0, 1, \ldots \tag{7.136}$$

For solving the fixed point problem $\mathbf{x} = \mathcal{S}(\mathbf{x})$ one can associate to it the iterative method

$$\mathbf{x}^{k+1} = F\left(\mathbf{x}^k, \mathcal{S}(\mathbf{x}^k), \ldots, \mathcal{S}^m(\mathbf{x}^k)\right), \quad n = 0, 1, \ldots \tag{7.137}$$

where $\mathcal{S}^{m+1}(\mathbf{x}) = \mathcal{S} \circ \mathcal{S}^m(\mathbf{x})$ and $\mathcal{S}^0(\mathbf{x}) = \mathbf{x}$. This approach is called full cycling or simply cycling. Conversely to any fixed point iteration of this form, one can associate a sequence transformation of the previous form. See Box 7.8 for the general algorithm, excluding the extrapolation method.

Box 7.8: Timestep n of vector extrapolation with cycling.

(i) Choose integers $n \geq 0$ and $k \leq 1$ and an initial vector $\mathbf{x}^k = \mathbf{x}_0^*$.

(ii) Compute $\mathbf{x}_1^*, \mathbf{x}_2^*, \ldots, \mathbf{x}_{n+k+1}^*$ via $\mathbf{x}_{m+1}^* = \mathcal{S}(\mathbf{x}_m^*)$.

(iii) Apply any of the four extrapolation methods, namely, MPE, RRE, MMPE, and SVD-MPE, to $\mathbf{x}_n^*, \mathbf{x}_{n+1}^*, \ldots, \mathbf{x}_{n+k+1}^*$, obtaining $\mathbf{s}_{k,n}$.

(iv) If $\mathbf{s}_{n,k}$ satisfies the accuracy test, stop.
Otherwise, set $\mathbf{x}_0^* = \mathbf{s}_{n,k}$ and go to Step (ii).

There is a variety of vector extrapolation methods, where the major two categories are polynomial methods and methods based on the ϵ-algorithm (**??**). In this presentation, only the first category is considered since the second requires a relatively large number of function evaluations per iteration, making it unsuitable for the present use-case (**?**).

? presents four different polynomial extrapolation methods. They all attempt to exprress the limit of the vector sequence as a linear combination of the p previous iterates, as follows

$$\mathbf{s} \approx \mathbf{s}_{k,m} = \sum_{j=0}^{m} \gamma_j \mathbf{w}^{k+j}, \tag{7.138}$$

where $\mathbf{s}$ is the limit of the vector sequence. The methods to be presented next appear naturally when considering the vector sequence generated by

$$\mathbf{w}^{k+1} = \mathbf{T}\mathbf{w}^k + \mathbf{d}, \tag{7.139}$$

where $\mathbf{I} - \mathbf{T}$ is non-singular. It is tighly connected to the solution of linear systems of equations. Considering the minimal polynomial of $\mathbf{T}$ with respect to $\Delta\mathbf{w}^k = \mathbf{w}^{k+1} - \mathbf{w}^k$

and $\boldsymbol{\epsilon}^k = \mathbf{w}^k - \mathbf{s}$[2], $P(\lambda)$,

$$P(\lambda) = \sum_{j=0}^{i} c_j \lambda^j, \quad c_i = 1, \tag{7.140}$$

where i is the degree of the polynomial, the limit of the sequence can be found exactly as

$$\mathbf{s} = \frac{\sum_{j=0}^{i} c_j \mathbf{w}^{k+j}}{\sum_{j=0}^{i} c_j}. \tag{7.141}$$

This can be derived considering the definition of $P(\lambda)$, $P(\mathbf{T})\boldsymbol{\epsilon}^k = 0$. Therefore,

$$\mathbf{0} = P(\mathbf{T})\boldsymbol{\epsilon}^k = \sum_{j=0}^{i} c_j \mathbf{T}^j \boldsymbol{\epsilon}^j = \sum_{j=0}^{i} c_j \boldsymbol{\epsilon}^{k+j}. \tag{7.142}$$

and so

$$0 = \sum_{i=0}^{k} c_i \boldsymbol{\epsilon}_{n+i} = \sum_{i=0}^{k} c_i \boldsymbol{x}_{n+i} - \left(\sum_{i=0}^{k} c_i \right) \mathbf{s}. \tag{7.143}$$

Solving this for $\mathbf{s}$, one obtains the desired result, provided $\sum_{j=0}^{i} c_j \neq 0$.

The coefficients of $P(\lambda)$ can be computed considering

$$\mathscr{W}^{i-1}\mathbf{c}' = -\Delta\mathbf{w}_{k+i}, \quad \mathbf{c}' = \left[c_0, c_1, \ldots, c_{i-1}\right]^T, \tag{7.144}$$

where $\mathscr{W}^i = [\Delta\mathbf{w}^k, \ldots, \Delta\mathbf{w}^{k+i}]$, since from the defintion of $P(\lambda)$, one has

$$\mathbf{0} = P(\mathbf{T})\Delta\mathbf{w}^k = \sum_{j=0}^{i} c_i \mathbf{T}^j \Delta\mathbf{w}^k = \sum_{j=0}^{i} c_i \Delta\mathbf{w}^{k+j}. \tag{7.145}$$

The degree of $P(\lambda)$ can be as large as the dimension of $\mathbf{w}$. Hence, to be practical, the minimal polynomial extrapolation (MPE), the reduced rank extrapolation (RRE), the modified minimal extrapolation (MMPE), and the single-value decomposition, minimal polynomial extrapolation (SVD-MPE) all choose a polynomial of a lesser degree. The approximations corresponding to each extrapolation method are presented in what follows.

MPE Solve the overdetermined linear system $\mathscr{W}^{m-1}\mathbf{c}' = -\Delta\mathbf{w}_{k+m}$ in the least-squares sense for $\mathbf{c}' = \left[c_0, c_1, \ldots, c_{m-1}\right]^T$. This amounts to solving the optimization problem

$$\min_{c_0, c_1, \ldots, c_{p-1}} \left\| \sum_{j=0}^{m-1} c_j \Delta\mathbf{w}^{k+j} + \Delta\mathbf{w}^{k+m} \right\|_2 \tag{7.146}$$

which can also be expressed as

$$\min_{\mathbf{c}'} \left\| \mathscr{W}^{m-1}\mathbf{c}' + \mathbf{w}^{k+m} \right\|_2, \quad \mathbf{c}' = \left[c_0, c_1, \ldots, c_{m-1}\right]^T. \tag{7.147}$$

With $c_0, c_1, \ldots, c_{k-1}$ available, set $c_m = 1$ and compute $\gamma_q = c_q / \sum_{j=0}^{m} c_j, q = 0, 1, \ldots, m$, provided $\sum_{j=0}^{m} c_j \neq 0$.

[2] A polynomial $P(\lambda)$ is said to be minimal with respect to a vector $\mathbf{a}$, if $P(\mathbf{T})\mathbf{a} = 0$ and it of least degree.

RRE Solve the overdetermined linear system $\mathscr{W}^m\boldsymbol{\gamma} = 0$ in the least-squares sense, subject to the constraint $\sum_{j=0}^{m}\gamma_j = 1$. This amounts to solving the optimization problem

$$\min_{\gamma_0,\gamma_1,\ldots,\gamma_m}\left\|\sum_{j=0}^{m}\gamma_j\Delta\mathbf{w}^{k+j}\right\| \text{ subject to } \sum_{j=0}^{m}\gamma_j = 1 \tag{7.148}$$

which can also be expressed as

$$\min_{\boldsymbol{\gamma}}\left\|\mathscr{W}^m\boldsymbol{\gamma}\right\|_2 \quad \text{subject to } \sum_{j=0}^{m}\gamma_j = 1; \quad \boldsymbol{\gamma} = \left[\gamma_0,\gamma_1,\ldots,\gamma_m\right]^T. \tag{7.149}$$

MMPE Consider a set of m linearly independent vectors $\mathbf{q}_j,\ j = 1,\ldots,m$. Solve the linear system

$$\left(\mathbf{q}^j,\mathscr{W}^{m-1}\mathbf{c}'\right) = -\left(\mathbf{q}^j,\Delta\mathbf{w}^{k+m}\right), \quad j = 1,\ldots,m, \tag{7.150}$$

which can also be expressed as

$$\sum_{j=0}^{m-1}\left(\mathbf{q}_j,\Delta\mathbf{w}^{k+j}\right)\mathbf{c}_j = -\left(\mathbf{q}^j,\Delta\mathbf{w}^{k+p}\right), \quad j = 1,\ldots,m. \tag{7.151}$$

This is, in fact, a system of m linear equations for the m unknowns $c_0,c_1,\ldots,c_{m-1}$. With $c_0,c_1,\ldots,c_{m-1}$ available, set $c_m = 1$ and compute $\gamma_q = c_q/\sum_{j=0}^{m}c_j, i = 0,1,\ldots,m$, provided $\sum_{j=0}^{m}c_j \neq 0$.

SVD-MPE Solve the standard l_2 constrained minimization problem

$$\min_{\mathbf{c}}\left\|\mathscr{W}^m\mathbf{c}\right\|_2 \quad \text{subject to } \|\mathbf{c}\|_2 = 1, \quad \mathbf{c} = \left[c_0,c_1,\ldots,c_m\right]^T. \tag{7.152}$$

The solution $\mathbf{c}$ is the right singular vector corresponding to the smallest singular value $\sigma_{\min}$ of $\mathscr{W}^m$, i.e., $\mathscr{W}^{m*}\mathscr{W}^m\mathbf{c} = \sigma_{\min}^2\mathbf{c}$, $\|\mathbf{c}\|_2 = 1$. It is assumed that $\sigma_{\min}$ is simple so that $\mathbf{c}$ is unique up to a multiplicative constant ϕ, $|\phi| = 1$.

With $c_0,c_1,\ldots,c_m$ available, compute $\gamma_q = c_q/\sum_{j=0}^{m}c_j, q = 0,1,\ldots,m$, provided $\sum_{j=0}^{m}c_j \neq 0$. The assumption that $\sigma_{\min}$ is simple guarantees the uniqueness of the γ_i.

When $m = 1$, MPE, RRE, MMPE, and SVD-MPE can be regarded as generalizations of the Aitken Δ^2-process to the vector case. Thus, when applied to the solution of a system of nonlinear equations using cycling

$$\mathbf{s}_{k,1} = \begin{cases} \mathbf{x}^k - \dfrac{\left(\Delta\mathbf{x}^k,\Delta\mathbf{x}^k\right)}{\left(\Delta\mathbf{x}^k,\Delta^2\mathbf{x}^k\right)}\Delta\mathbf{x}^k & \text{for MPE,} \\ \mathbf{x}^k - \dfrac{\left(\Delta^2\mathbf{x}^k,\Delta\mathbf{x}^k\right)}{\left(\Delta^2\mathbf{x}^k,\Delta^2\mathbf{x}^k\right)}\Delta\mathbf{x}^k & \text{for RRE,} \\ \mathbf{x}^k - \dfrac{\left(\mathbf{q}_1,\Delta\mathbf{x}^k\right)}{\left(\mathbf{q}_1,\Delta^2\mathbf{x}^k\right)}\Delta\mathbf{x}^k & \text{for MMPE,} \\ \mathbf{x}^k - \dfrac{\left(\mathbf{g}_0,\Delta\mathbf{x}^k\right)}{\left(\mathbf{g}_0,\Delta^2\mathbf{x}^k\right)}\Delta\mathbf{x}^k & \text{for SVD-MPE.} \end{cases} \tag{7.153}$$

? also suggest cycling with frozen γ_i, where after some iterations the γ_i are frozen and reused henceforth. A parallel version of the full cycling procedure is also described.

Connection to Krylov subspace methods According to **?**, the so-called Krylov subspace methods are closely related to the vector extrapolation methods presented above. When the latter is applied to vector sequences obtained using fixed-point iterative methods to nonsingular linear systems of equations, they are mathematically equivalent. More precisely, the MPE and the RRE methods are mathematically equivalent to the methods of Arnoldi and generalized minimal residual (GMR).

However, Krylov subspace methods and extrapolation methods differ in their algorithmic aspects entirely: The only input of the former is a procedure that performs the matrix-vector multiplication without explicitly knowing the matrix coefficient matrix. The latter takes as their only input a vector sequence that results from a fixed-point iterative scheme without knowing the matrix coefficient.

In **?**, a Krylov-subspace method is proposed in the context of FSI. However, as pointed out by **?**, the correct term for this approach should be instead a "Krylov-based vector extrapolation" method. The method proposed can be obtained by applying the RRE to the sequence of residuals computed as $\Delta\mathbf{r}_i^* = \mathbf{x}_i^* - \mathbf{x}^k$, where the subscript i concerns the internal loop of the vector extrapolation method, and whose limit is $\mathbf{0}$. **?** argues that these residual differences have unfavorable numerical properties and should be avoided.

7.6 Multipoint iteration functions with memory

Multipoint iteration functions are rarer and are not thoroughly investigated in this exposition. One can mention the Eirola-Nevanlinna family of methods (**?**) and Netwon-Krylov method that reuses the Krylov subspace from previous iterations (**?**).

7.7 Summary

The overview provided in this chapter centered on implicit techniques for partitioned multi-physics coupled applications. The problem can be slightly reformulated as a system of nonlinear equations, where the residual represents the difference between the input and output after the fixed-point procedure has been applied. This allows the problem to be solved using a variety of methods that have been described in the literature. The key issues in the current application are the memory requirements of the numerical technique due to a potential high number of unknowns and the number of residual evaluations, as each assessment necessitates a significant computational effort. Table 7.1 provides information on the memory requirements, the number of function evaluations per nonlinear iteration, and some pertinent observations for each technique under consideration.

Except for the MMPE and SVD-MPE approaches, all the methods discussed in the present chapter are analyzed in the next chapter (Chapter **??**). As seen in Table 7.2, where the update formulae for the methods are compiled, their implementation is made simpler by the similarities in the update formula for the value of the unknown variable at each nonlinear iteration.

Table 7.1: Summary of the comparison between method for the solution methods of non-linear systems of equations. n here denotes the number of unknowns and m denotes depending on the context the number of previous iterates considered, the number of fixed point evaluations or the size of the Krylov subspace.

Method	Memory requirements	Nr function evaluations per iteration	Observations
Fixed-point iteration	2 $(n \times 1)$ vectors	1	•Often diverges. •Simplest method. •Memory efficient.
Underrelaxation	2 $(n \times 1)$ vectors	1	•Simple. •Improved stability over fixed-point. •Need to manually choose a relaxation parameter.
Aitken relaxation	3 $(n \times 1)$ vectors	1	•Very popular in FSI. •Dynamic relaxation. •Improved stability over fixed-point.
Broyden-like family (**?**)	2 $(n \times 1)$ vectors 2 $(n \times (m-1))$ matrices	1	•$\mathcal{O}(nm^2)$ computation complexity (QR decomposition). •Low number of function evaluations •Superlinear convergence when $m = 1$.
Broyden's method (**?**)	Up to $(m+2)(n \times 1)$ matrices	1	•$\mathcal{O}(n)$ computation complexity. •Low storage. •Superlinear convergence.
Newton-Krylov	Up to $(m+1)$ $(n \times 1)$ vectors	$m+1^*$	•Large number of iterations possible. •Popular for the solution of systems of nonlinear equations. •Quadratic convergence under appropriate conditions.
Vector extrapolation in cycling mode	$(m+2)$ $(n \times 1)$ vectors	$m+1$	•Large number of function evaluations. •$\mathcal{O}(nm^2)$ computational complexity (QR decomposition).

* The number of function evaluations in the Newton-Krylov methods will depend on how many iterations it will take for the inner loop to converge. There is a function evaluation per iteration of the inner loop.

Page intentionally left blank.

Table 7.2: Summary of the update formulas for the solution methods of non-linear systems of equations.

Method	Update formula
Fixed-point	$\mathbf{x}^{k+1} = \mathbf{x}^k - \mathcal{R}^k$
Underrelaxation	$\mathbf{x}^{k+1} = \mathbf{x}^k - \omega\mathcal{R}^k,$ $0 < \omega < 1.$
Aitken relaxation	$\mathbf{x}^{k+1} = \mathbf{x}^k - \omega^{(k)}\mathcal{R}^k,$ $\omega^{(k)} = -\omega^{(k-1)} \frac{\left(\mathbf{r}^{(k)} - \mathbf{r}^{(k-1)}\right)^{\mathrm{T}} \mathbf{r}^{(k-1)}}{\left(\mathbf{r}^{(k)} - \mathbf{r}^{(k-1)}\right)^2}.$
Broyden's family	$\mathbf{x}^{k+1} = \mathbf{x}^k - \left(G_{\mathcal{R}}^{k-m} + \left(\mathscr{X}^k - G_{\mathcal{R}}^{k-m}\mathscr{R}^k\right)\mathbf{V}^{k\,T}\right)\mathcal{R}^k,$ $\mathbf{V}^{k\,T} = \mathbf{M}^{k-1}\mathbf{N}^{k\,T},$ Type I: $M^k = \mathscr{X}^{k\,T} G_{\mathcal{R}}^k \mathscr{R}^k, \quad N^{k\,T} = \mathscr{X}^{k\,T} G_{\mathcal{R}}^k,$ Type II: $M^k = \mathscr{R}^{k\,T}\mathscr{R}^k, \quad N^{k\,T} = \mathscr{R}^{k\,T}$ $\mathbf{V}^k$ is usually computed using QR-decomposition.
Anderson's family	$\mathbf{x}^{k+1} = \mathbf{x}^k - \left(-\beta\mathbf{I} + \left(\mathscr{X}^k + \beta\mathscr{R}^k\right)\mathbf{V}^{k\,T}\right)\mathcal{R}^k,$ $\mathbf{V}^{k\,T} = \mathbf{M}^{k-1}\mathbf{N}^{k\,T},$ Type I: $M^k = \mathscr{X}^{k\,T}\mathscr{R}^k, \quad N^{k\,T} = \mathscr{X}^{k\,T},$ Type II: $M^k = \mathscr{R}^{k\,T}\mathscr{R}^k, \quad N^{k\,T} = \mathscr{R}^{k\,T}$ $\mathbf{V}^k$ is usually computed using QR-decomposition.
Newton-Krylov	$\mathbf{x}^{k+1} = \mathbf{x}^k + \Delta\mathbf{x}^k$ $J_{\mathcal{R}}(\mathbf{x}^k)\Delta x^k = -\mathcal{R}^k$ solved using a Krylov method to accuracy $\|J_{\mathcal{R}}(\mathbf{x}^k)\Delta x_m^* + \mathcal{R}(\mathbf{x}^k)\| \leq \eta\|\mathcal{R}(\mathbf{x}^k)\|$
Vector extrapolation in cycling mode	$\mathbf{x}^{k+1} = \sum_{j=0}^m \gamma_j \mathcal{S}^{i+j}(\mathbf{x}^k),$ $\gamma_q = c_q / \sum_j c_j, \quad q = 0, 1, \ldots, m,$ **MPE**: $\mathbf{c}' = [c_0, c_1, \ldots, c_{m-1}]^T = \arg\min_{\bar{\mathbf{c}}} \left\| \mathscr{S}^{m-1}\bar{\mathbf{c}}' + \mathcal{S}^{i+m}(\mathbf{x}^k) \right\|_2,$ $c_m = 1,$ **RRE**: $\boldsymbol{\gamma} = [\gamma_0, \gamma_1, \ldots, \gamma_m]^T = \arg\min_{\bar{\boldsymbol{\gamma}}} \left\| \mathscr{S}^m \hat{\boldsymbol{\gamma}} \right\|_2,$ Subject to $\sum_{j=0}^m \gamma_j = 1.$ **MMPE**: $\left(\mathbf{q}^j, \mathscr{S}^{m-1}\mathbf{c}'\right) = -\left(\mathbf{q}^j, \mathcal{S}^{i+m}(\mathbf{x}^k)\right), \quad j = 1, \ldots, m,$ for a set of m linearly independent vectors $\mathbf{q}_j$, $j = 1, \ldots, m.$ $\mathbf{c}' = [c_0, c_1, \ldots, c_{m-1}]^T.$ **SVD-MPE**: $\mathbf{c} = \arg\min_{\bar{\mathbf{c}}} \left\| \mathscr{S}^m \bar{\mathbf{c}} \right\|_2 \quad$ subject to $\|\bar{\mathbf{c}}\|_2 = 1,$ $\mathbf{c} = \left[c_0, c_1, \ldots, c_m\right]^T.$

Chapter 8

Numerical results for the implicit coupling schemes

This chapter provides validation and comparison results for most of the implicit solution schemes discussed in Chapter 7. The methods considered are the fixed-point method (see Section 7.3.1), the underrelaxation method (see Section 7.3.3), the Aitken relaxation (see Section 7.4.1), the Broyden-like family of methods, including Broyden's method (see Section 7.4.2), the Newton-GMRES method (see Section 7.5.2) and the polynomial vector extrapolation methods, MPE and RRE, in cycling mode (see Section 7.5.3). The use of predictors is also considered (see Section 7.2.1) as a way to improve efficiency. It focuses on two thermomechanical examples for which reference solutions are available in the literature: the quasi-static finite strain thermo-elastic expansion of an infinitely long-thick walled cylinder and the necking of a thermo-elastoplastic circular bar.

8.1 Expansion of a thermoelastic thick-walled cylinder

The following numerical example concerns the quasi-static finite strain thermo-elastic expansion of an infinitely long thick-walled cylinder, as presented in **?**. **?** and **?** also present results regarding this problem, although the dimensions of the cylinder considered there are different from the ones used in the current work. The example is comprised of an infinitely long cylinder with an inner radius of $r_0 = 5\,\text{mm}$ and an outer radius of $r_1 = 15\,\text{mm}$. A displacement-driven problem is produced by enforcing an increasing radial displacement at the inner radius with a constant rate of $\dot{u}_0$. Following **?**, the maximum displacement is set to 10 mm, which clearly involves large deformations. Zero heat flux is imposed at the inner radius, whereas the temperature at the outer radius is set to a reference temperature T_0. The initial temperature of the cylinder is also T_0. A sketch of the problem, including the initial and boundary conditions described, is shown in Figure 8.1. A decoupled Neo-Hookean free energy function characterizes the thermo-elastic material considered in this analysis, following **?**. Table 8.1 contains all material and model properties of relevance.

A plane strain analysis is used to solve the problem, implying a null displacement and heat flux in the axial direction. Linear quadratic elements (QUAD4) are used in the FEM discretization employed. Thanks to the axial symmetry of the problem, only one-

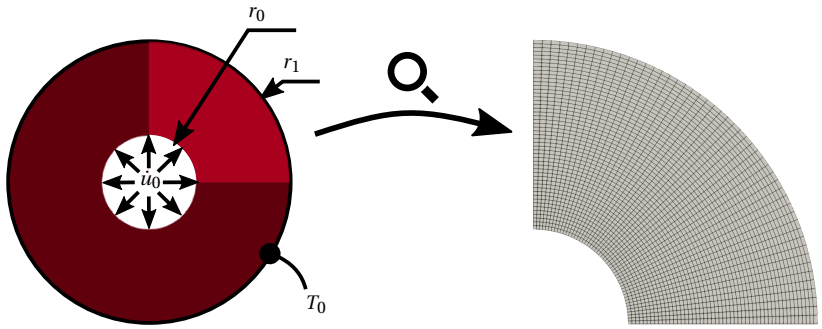


Figure 8.1: Initial and boundary conditions considered in the quasi-static finite strain thermo-elastic expansion of an infinitely long thick-walled cylinder, and corresponding FEM mesh (QUAD4) used.

Table 8.1: Material properties and initial and boundary conditions for the problem concerning the quasi-static finite strain thermo-elastic expansion of an infinitely long thick-walled cylinder.

Material Properties			Effective value
Density	ρ	($\mathrm{Ns^2mm^{-4}}$)	7.8×10^{-9}
Bulk modulus	κ	($\mathrm{Nmm^{-2}}$)	164.206×10^3
Shear modulus	μ	($\mathrm{Nmm^{-2}}$)	80.140×10^3
Conductivity	k	($\mathrm{Ns^{-1}K^{-1}}$)	45
Heat capacity	C_V	($\mathrm{mm^2s^{-2}K^{-1}}$)	460×10^6
Coefficient of thermal expansion	α_T	($\mathrm{K^{-1}}$)	0 - 150×10^{-6}
Boundary Conditions			
Inner radius	r_0	(mm)	5
Outer radius	r_1	(mm)	15
Inner radius rate of displacement	$\dot{u}_0$	($\mathrm{mms^{-1}}$)	0.1; 0.25; 0.5
Heat at inner radius	q_1	($\mathrm{Ns^{-1}mm^{-1}}$)	0
Temperature outer radius	T_1	(K)	273.15
Initial Conditions			
Initial temperature	T_0	(K)	273.15
Reference value			
Temperature at inner radius ($r = r_0$)	(K)		

fourth of the cylinder is considered, as depicted in Figure 8.1. Except when explicitly indicated, the mesh employed contains 2601 nodes and 2500 elements.

The displacement is imposed in 200 equal time steps $\Delta t = 0.1\,\mathrm{s}$, A quasi-static solution is computed for the mechanical problem using a backward Euler integration scheme. The transient temperature field is integrated in time employing the generalised-α method with $\rho_{\infty,T} = 1.0$.

8.1.1 Validation of the Numerical Results

To validate the results presented in this work regarding the quasi-static finite strain thermo-elastic expansion of an infinitely long thick-walled cylinder, the problem is solved considering an $\alpha_T = 1.65 \times 10^{-4}\,\mathrm{K}^{-1}$ and a $\alpha_T = 1.65 \times 10^{-5}\,\mathrm{K}^{-1}$, at three different displacement rates for the inner radius, ($\dot{u}_0 = 0.1\,\mathrm{mms}^{-1}$, $0.25\,\mathrm{mms}^{-1}$, and $0.5\,\mathrm{mms}^{-1}$). Reference results for this problem configuration are available in **?** and provide a suitable comparison for validating the results obtained in the present work. There is however one caveat, since in **?** the results reported are supposedly for coefficients of thermal expansion equal to $10^{-5}\,\mathrm{K}^{-1}$ and $10^{-4}\,\mathrm{K}^{-1}$. The values used in the present work lead to the best fit between the present and the reference results. The values chosen for the expansion coefficient lead to weak and strong coupling. Using the latter value for the thermal expansion coefficient leads to instabilities when using a fixed-point explicit or implicit scheme with the isothermic split (**??**).

The temperature evolutions as a function of inner raidus displacement for a point located at the inner radius of the cylinder are presented in Figures 8.2 and 8.3 chosing $\alpha_T = 1.65 \times 10^{-5}\,\mathrm{K}^{-1}$ and $\alpha_T = 1.65 \times 10^{-4}\,\mathrm{K}^{-1}$, respectively. Figure 8.4 presents the temperature distribution in half a transversal section of the thick-walled cylinder at increasing inner radius displacements. A good agreement between the present work's numerical results and those found in **?** can be observed for all displacement rates considered. The temperature at the inner radius suffers a sharp decrease and then, as the displacement increases, tends to the reference temperature T_0. This decrease is due to the thermo-elastic heating effect and the fact that heat cannot be supplied from the outer radius fast enough to prevent this drop in temperature near the inner radius.

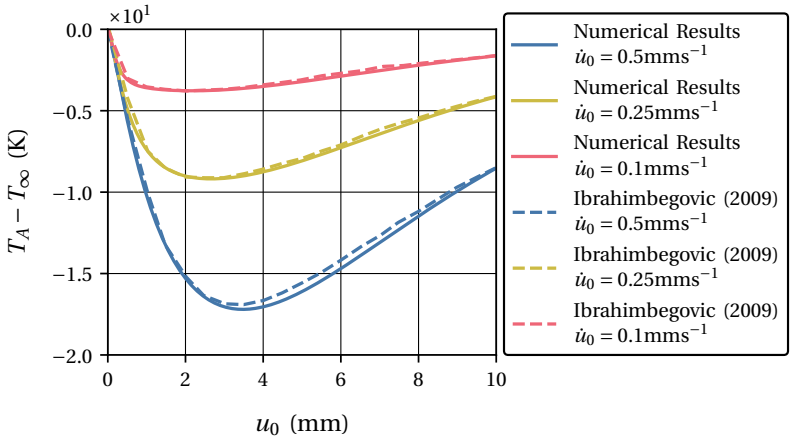


Figure 8.2: Difference between the temperature at the inner radius and the reference temperature for the expansion of the thick-walled cylinder with $\alpha_T = 1.65 \times 10^{-5}\,\text{K}^{-1}$ and at different displacement rates for the inner radius ($\dot{u}_0 = 0.1\,\text{mms}^{-1}$, $0.25\,\text{mms}^{-1}$, and $0.5\,\text{mms}^{-1}$

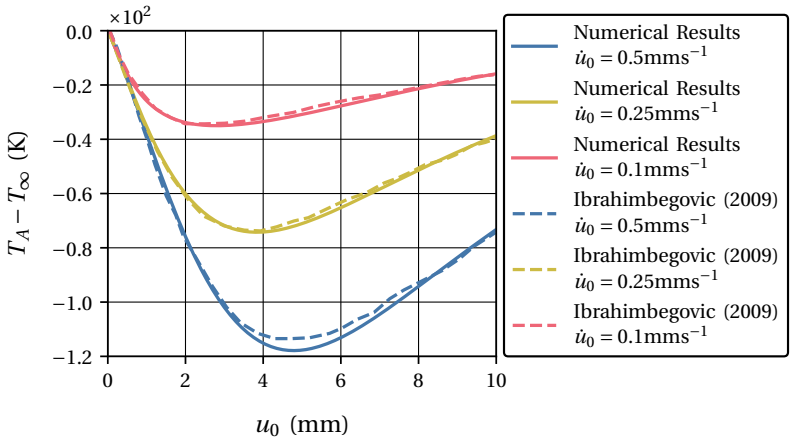


Figure 8.3: Difference between the temperature at the inner radius and the reference temperature for the expansion of the thick-walled cylinder with $\alpha_T = 1.65 \times 10^{-4}\,\text{K}^{-1}$ and at different displacement rates for the inner radius ($\dot{u}_0 = 0.1\,\text{mms}^{-1}$, $0.25\,\text{mms}^{-1}$, and $0.5\,\text{mms}^{-1}$

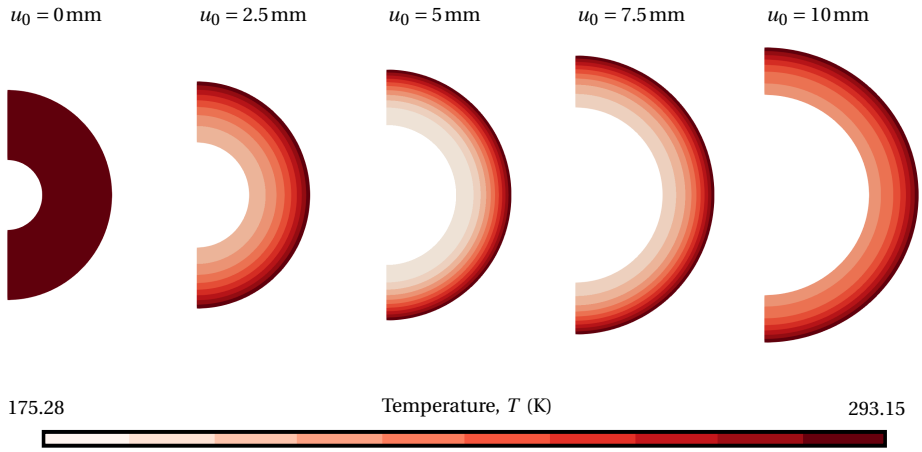


Figure 8.4: Temperature distribution in half a transversal section of the thick-walled cylinder at increasing inner radius displacements for $\alpha_T = 1.65 \times 10^{-4}\,\mathrm{K}^{-1}$ and $\dot{u}_0 = 0.5\,\mathrm{mms}^{-1}$.

8.1.2 Evaluation and comparison of implicit solution methods for the coupled problem

The following contains the results concerning the evaluation and comparison of the implicit solution methods for coupled problems considered in this work. The discussion starts with the methods that require only one evaluation of the residual per nonlinear iteration. The Broyden-like family of methods fall into this category too but are considered by themselves and are presented next. The Newton-Krylov methods are also analyzed, followed by the vector extrapolation methods in cycling mode. The discussion ends with the comparison between the best methods in each class and the effect of predictors on the efficiency of the implicit methods.

The analysis presented is based mainly on three pieces of information. The first concerns how the residual evolves as a function of the nonlinear iterations. The second is the number of function evaluations needed to solve the coupled problem to the desired accuracy at each time step. The third is the total number of residual evaluations needed to solve the coupled problem to the desired accuracy as a function of the thermal expansion coefficient. The larger the thermal expansion coefficient, the stronger the coupling between the thermal and mechanical fields, and the more challenging the problem is to solve. The evaluation of the residual implies the solution of the thermal and mechanical problems one after the other and takes the lion's share of computational time. Regarding conclusions related to efficiency, it suffices when comparing methods of the same class to consider the number of residual evaluations. The computational time is only analyzed in the comparison between the best methods of each class.

Only the displacement rate $\dot{u}_0 = 0.5\,\mathrm{mms}^{-1}$ is considered, with the thermal expansion coefficient varying from $0\,\mathrm{K}^{-1}$ to $1.5 \times 10^{-4}\,\mathrm{K}^{-1}$.

8.1.2.1 Methods with only one residual evaluation per iteration

The methods with only one residual evaluation per nonlinear iteration considered are the fixed-point method (see Section 7.3.1), the underrelaxation method (see Section 7.3.3), the Aitken relaxation (see Section 7.4.1) and Broyden's method, Type I and II (see Section 7.4.2). These are, a priori, the most economical methods regarding residual evaluations, as only one is performed per nonlinear iteration. The underrelaxation is performed with $\omega = 0.5$, and the first relaxation coefficient for the Aitken relaxation is also set to 0.5.

Figure 8.5 presents the residual in percentage as a function of the number of nonlinear iterations in the first time step with $\alpha_T = 1.5 \times 10^{-4}\,\mathrm{K}^{-1}$ and $\dot{u}_0 = 0.5\,\mathrm{mms}^{-1}$. As reported by **?**, the fixed-point scheme cannot converge for this thermal expansion coefficient value. The other methods converge approximately linearly, with the Broyden and Aitken relaxation methods taking the same number of nonlinear iterations/residual evaluations. Furthermore, the two Broyden methods are almost visually indistinguishable. The underrelaxation methods take more iterations to converge, which could be improved by tuning the relaxation coefficient further.

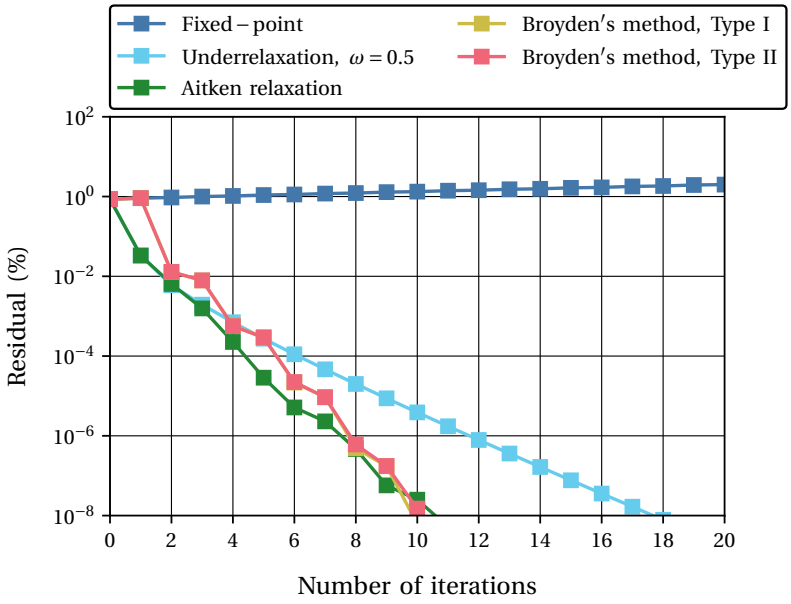


Figure 8.5: Residual in percentage as a function of the number of nonlinear iterations in the first time step for the implicit methods that perform only one evaluation per nonlinear iteration in the solution of the quasi-static expansion of a thermoelastic thick-walled cylinder with $\alpha_T = 1.5 \times 10^{-4}\,\mathrm{K}^{-1}$ and $\dot{u}_0 = 0.5\,\mathrm{mms}^{-1}$.

Figure 8.6 presents the number of nonlinear iterations/residual evaluations needed to solve the coupled thermomechanical problem at each time step and the total (cumulative) number of iterations needed. The strength of the coupling, tightly connected to the difficulty in solving the coupled problem and hence the number of nonlinear iterations needed to solve it, seems to be approximately uniform across the displacement range considered, with a slight decrease as the displacement increases. The Broyden methods show a very similar behavior, besting the Aitken relaxation at

each time step by at most one or two nonlinear iterations. This leads to a sizable difference in the total number of function evaluations needed to solve the problem from start to finish. The underrelaxation method underperforms again, showing even more difficulty in solving the problem as the displacement increases.

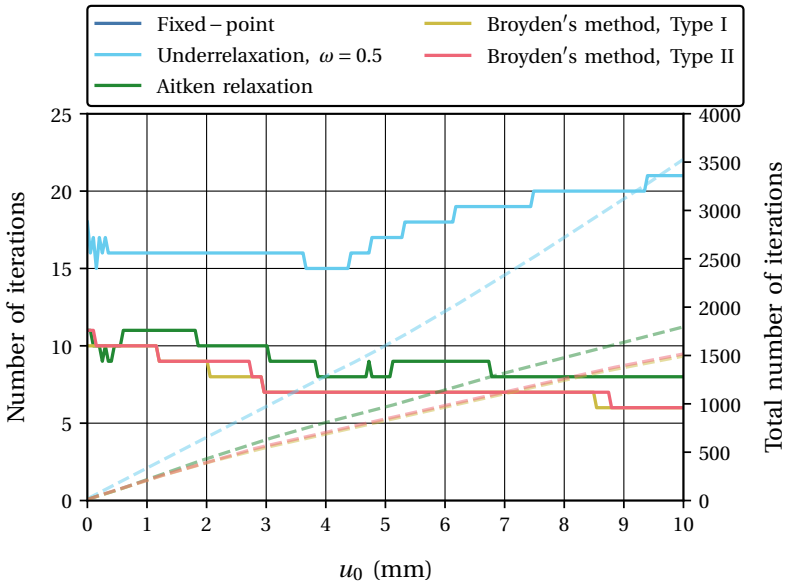


Figure 8.6: number of nonlinear iterations needed to solve the coupled problem at each time step and the total number of iterations needed to solve the coupled problem for the implicit methods that perform only one evaluation per nonlinear iteration in the solution of the quasi-static expansion of a thermoelastic thick-walled cylinder with $\alpha_T = 1.5 \times 10^{-4}\,\mathrm{K}^{-1}$ and $\dot{u}_0 = 0.5\,\mathrm{mm\,s}^{-1}$.

Figure 8.7 presents the total number of residual evaluations, in this case, corresponding also to the total number of nonlinear iterations, as a function of the thermal expansion coefficient, which controls the strength of the coupling between the thermal and the mechanical fields, respectively. The residual evaluations correlate strongly with the total CPU time, accounting for the most significant portion of the computational time. The most efficient methods are the two Broyden methods, followed by the Aitken relaxation. The underrelaxation method performs poorly throughout the range of values considered for the thermal expansion coefficient, and the fixed-point becomes increasingly slow as the coupling gets stronger, eventually failing to converge for $\alpha_T = 1.5 \times 10^{-4}\,\mathrm{K}^{-1}$.

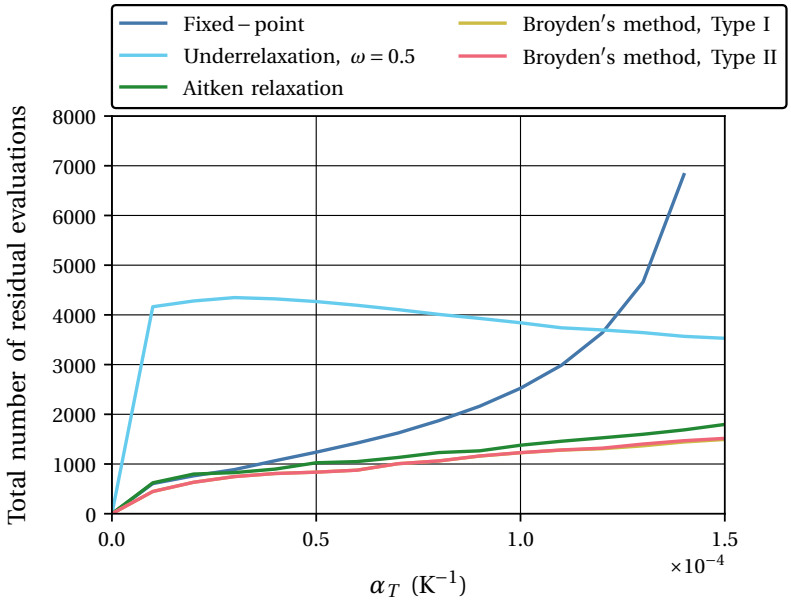


Figure 8.7: Total number of residual evaluations as a function of the thermal expansion coefficient for the implicit methods that perform only one evaluation per nonlinear iteration in the solution of the quasi-static expansion of a thermoelastic thick-walled cylinder with $\alpha_T = 0\,K^{-1}$ to $1.5 \times 10^{-4}\,K^{-1}$ and $\dot{u}_0 = 0.5\,mms^{-1}$.

8.1.2.2 Broyden-like method

The Broyden-like methods considered (see Section 7.4.2) employ as group sizes s = 1, 2, 3, 6 with a maximum number of previous iterations available equal to 6. The mixing parameters considered are $\beta = -1$, 2×10^{-3}, 2×10^{-2}. Their choice is based on prior tuning, which was not exhaustive. All combinations are also considered with both Type I and Type II updating for the approximation to the Jacobian.

Figures 8.8 and 8.9 present the residual in percentage as a function of the number of nonlinear iterations in the first time step with $\alpha_T = 1.5 \times 10^{-4}\,K^{-1}$ and $\dot{u}_0 = 0.5\,mms^{-1}$ for Broyden-like methods with Type I and Type II updates, respectively. There is no marked difference between methods that employ a Type I and Type II update. However, the choice of the mixing parameter and the group size leads to different behaviors for the residual. For $\beta = -1$, the methods using s = 1, 2, and 3 behave much the same with an approximately linear convergence rate and need the fewest iterations to reach the desired accuracy. For $s = 6$, the residual behaves differently, plateauing for several nonlinear iterations. Despite this, it only takes a few more iterations to converge when compared with the methods employing s = 1, 2, and 3. When $\beta = 2 \times 10^{-3}$ and 2×10^{-2} are utilized, they exihibt a similar trend, with residual plateuas for a few iterations before decreasing. This behavior can partly be explained by the implementation, which follows the suggestion found in **?**. Due to memory concerns, the approximation to the inverse of the Jacobian is only updated after the next group of residual evaluations has been filled (see Section 7.4.2). This approach can lead to a momentaneous poor approximation to the inverse of the Jacobian and thus stagnation of the residual. The choice of $\beta = 2 \times 10^{-2}$ seems to lead to slightly fewer iterations before the desired

accuracy is reached.

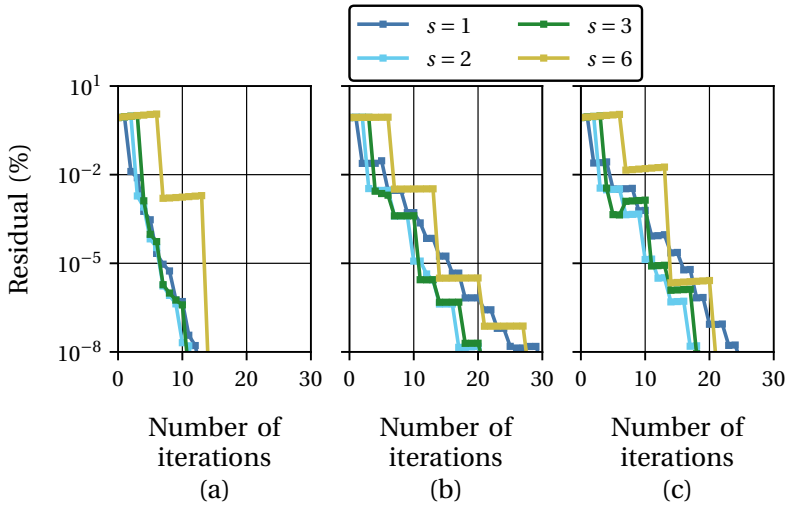


Figure 8.8: Residual in percentage as a function of the number of nonlinear iterations in the first time step for Broyden-like methods with Type I update and group sizes $s = 1$, 2, 4 and 6: (a) $\beta = -1$, (b) $\beta = 2 \times 10^{-3}$, and (c) $\beta = 2 \times 10^{-2}$ in the solution of the quasi-static expansion of a thermoelastic thick-walled cylinder with $\alpha_T = 1.5 \times 10^{-4}\,\mathrm{K}^{-1}$ and $\dot{u}_0 = 0.5\,\mathrm{mms}^{-1}$.

Figures 8.10 and 8.11 present the number of nonlinear iterations/number of function evaluations needed to solve the coupled problem at each time step and the total (cumulative) number of iterations needed. Again the differences between the Broyden-like methods using Type I and Type II updates are not pronounced. Perhaps the most noticeable difference is for $\beta = 2 \times 10^{-3}$ and 2×10^{-2}, where the total number of iterations needed to solve the coupled thermomechanical problem oscillates more strongly between time steps for the Type I methods. For $\beta = -1$, the methods using group sizes of 1, 2, and 3 display a higher efficiency, which disappears as the displacement increases. For $\beta = 2 \times 10^{-2}$ and 2×10^{-3}, the results seem to indicate the need for fewer iterations when using group sizes of 2 and 3.

Figures 8.12 presents the total number of residual evaluations as a function of the thermal expansion coefficient, respectively, for the Broyden-like methods with Type I update considered. The same results are presented in Figures 8.13 for the Broyden-like methos with Type II update. The most efficient methods employ a mixing parameter equal to $\beta = -1$ and group sizes equal to 1, 2, and 3. The choices of 2×10^{-3} and 2×10^{-2} for the mixing parameter lead to less efficient methods, which need a more significant amount of residual evaluations and thus require more CPU time to solve the coupled thermomechanical problem completely.

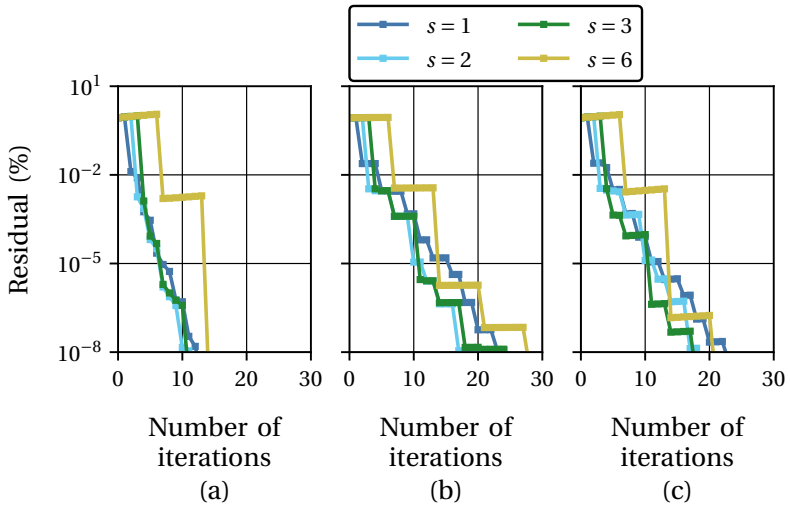


Figure 8.9: Residual in percentage as a function of the number of nonlinear iterations in the first time step for Broyden-like methods with Type II update and group sizes $s = 1$, 2, 4 and 6: (a) $\beta = -1$, (b) $\beta = 2 \times 10^{-3}$, and (c) $\beta = 2 \times 10^{-2}$ in the solution of the quasi-static expansion of a thermoelastic thick-walled cylinder with $\alpha_T = 1.5 \times 10^{-4}\,\mathrm{K}^{-1}$ and $\dot{u}_0 = 0.5\,\mathrm{mms}^{-1}$.

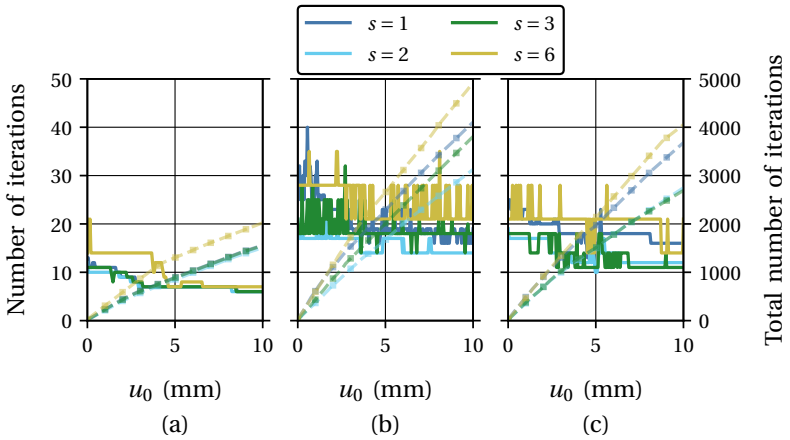


Figure 8.10: Number of nonlinear iterations needed to solve the coupled problem at each time step and the total number of iterations needed to solve the coupled problem for Broyden-like methods with Type I update and group sizes $s = 1$, 2, 4 and 6: (a) $\beta = -1$, (b) $\beta = 2 \times 10^{-3}$, and (c) $\beta = 2 \times 10^{-2}$ in the solution of the quasi-static expansion of a thermoelastic thick-walled cylinder with $\alpha_T = 1.5 \times 10^{-4}\,\mathrm{K}^{-1}$ and $\dot{u}_0 = 0.5\,\mathrm{mms}^{-1}$.

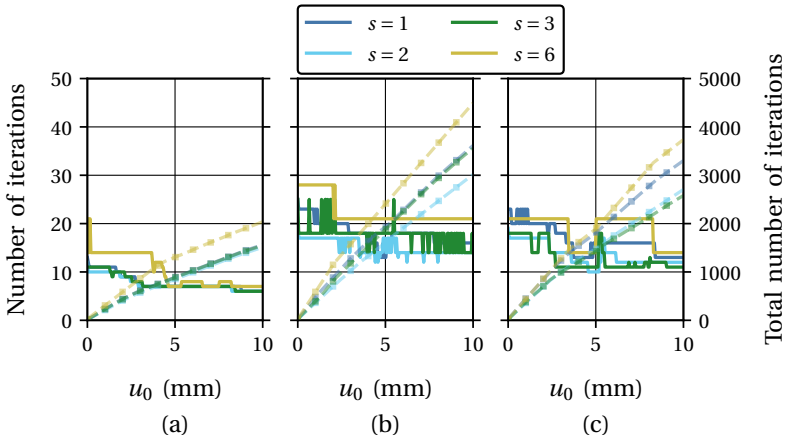


Figure 8.11: Number of nonlinear iterations needed to solve the coupled problem at each time step and the total number of iterations needed to solve the coupled problem for Broyden-like methods with Type II update and group sizes $s = 1$, 2, 4 and 6: (a) $\beta = -1$, (b) $\beta = 2\times10^{-3}$, and (c) $\beta = 2\times10^{-2}$ in the solution of the quasi-static expansion of a thermoelastic thick-walled cylinder with $\alpha_T = 1.5\times10^{-4}\,\mathrm{K}^{-1}$ and $\dot{u}_0 = 0.5\,\mathrm{mms}^{-1}$.

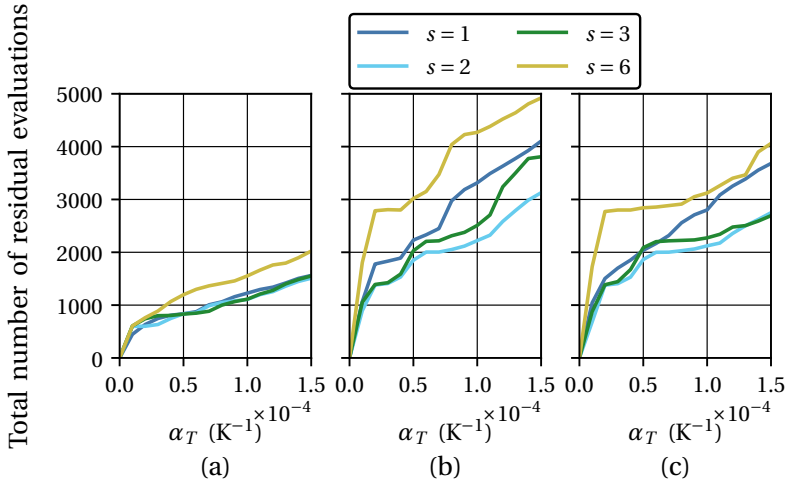


Figure 8.12: Total number of residual evaluations as a function of the thermal expansion coefficient for the implicit methods for Broyden-like methods with Type I update and group sizes $s = 1$, 2, 4 and 6: (a) $\beta = -1$, (b) $\beta = 2\times10^{-3}$, and (c) $\beta = 2\times10^{-2}$ in the solution of the quasi-static expansion of a thermoelastic thick-walled cylinder with $\alpha_T = 0\,\mathrm{K}^{-1}$ to $1.5\times10^{-4}\,\mathrm{K}^{-1}$ and $\dot{u}_0 = 0.5\,\mathrm{mms}^{-1}$.

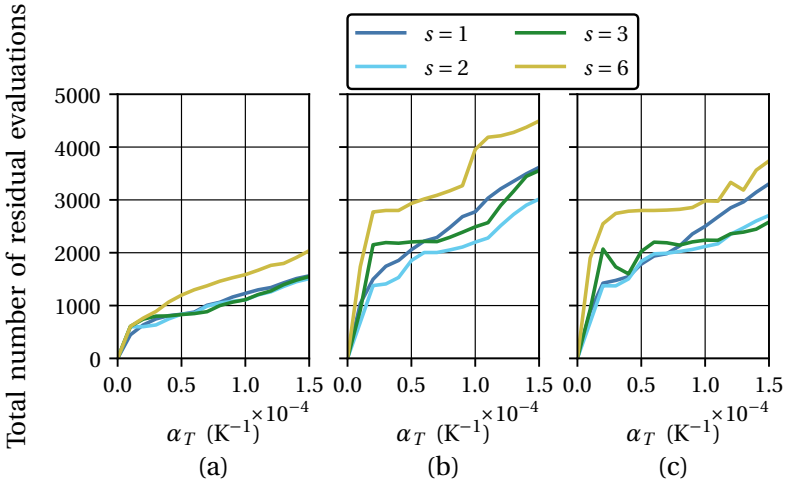


Figure 8.13: Total number of residual evaluations as a function of the thermal expansion coefficient for Broyden-like methods with Type II update and group sizes $s = 1$, 2, 4 and 6: (a) $\beta = -1$, (b) $\beta = 2 \times 10^{-3}$, and (c) $\beta = 2 \times 10^{-2}$ in the solution of the quasi-static expansion of a thermoelastic thick-walled cylinder with $\alpha_T = 0\,\mathrm{K}^{-1}$ to $1.5 \times 10^{-4}\,\mathrm{K}^{-1}$ and $\dot{u}_0 = 0.5\,\mathrm{mms}^{-1}$.

8.1.2.3 Newton-GMRES method

The Newton-Krylov method examined here uses as the Krylov subspace solver the GMRES method (see Section 7.5.2). Different forcing terms are employed. The values considered are $\eta = 10^{-1}$, 10^{-3} and 10^{-5}. The Eisenstat-Walker scheme for the adaptive choice of the forcing term is also utilized.

Figure 8.14 presents the residual in percentage as a function of the number of nonlinear iterations in the first time step with $\alpha_T = 1.5 \times 10^{-4}\,\mathrm{K}^{-1}$ and $\dot{u}_0 = 0.5\,\mathrm{mms}^{-1}$. For $\eta = 10^{-5}$, the Newton-GMRES method converges quadratically since the Newton system is solved to a finer accuracy, thus closely approximating the Newton-Raphson scheme. As η increases, the convergence rate as a function of the number of nonlinear iterations slows down. The method employing the Eisenstat-Walker scheme performs similarly to the method using a constant forcing term equal to 10^{-1}. Keep in mind that within each nonlinear iteration, the Newton-Krylov methods may evaluate the function several times, such that converging in fewer nonlinear iterations does not necessarily imply a more efficient method regarding computational time spent solving the coupled thermomechanical problem.

Figure 8.15 presents the number of nonlinear iterations needed to solve the coupled problem at each time step and the total (cumulative) number of iterations needed. The strength of the coupling seems to be approximately uniform across the displacement range considered, as the number of iterations taken by each method remains approximately constant as the displacement increases. As hinted by the results already discussed regarding the residual as a function of the nonlinear iterations in the first time step, the smaller the forcing, the fewer the number of nonlinear iterations needed to solve the coupled problem at each time step. The method that uses the Eisenstat-Walker scheme performs similarly to the method using

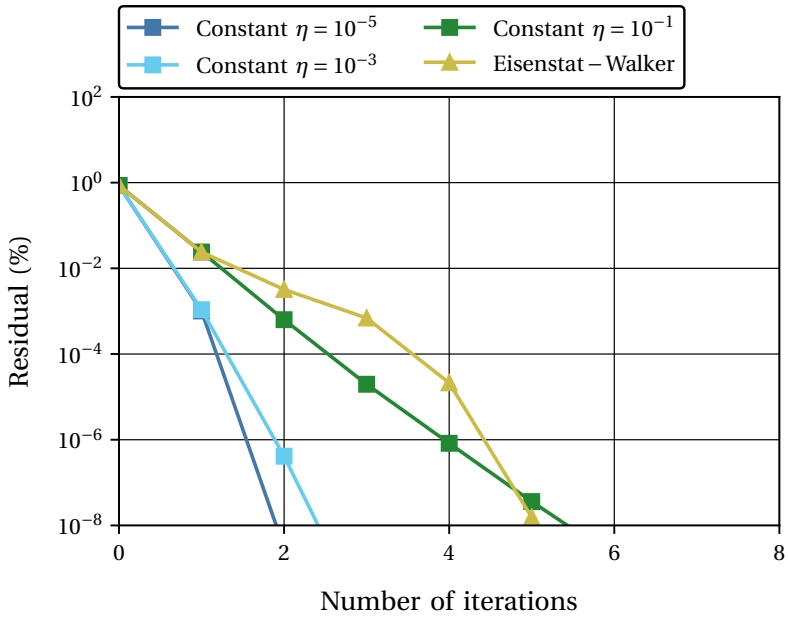


Figure 8.14: Residual in percentage as a function of the number of nonlinear iterations in the first time step for the Newton-GMRES method with a constant forcing term ($\eta = 10^{-5},\ 10^{-3},\ 10^{-1}$) and the Eisenstat-Walker scheme in the solution of the quasi-static expansion of a thermoelastic thick-walled cylinder with $\alpha_T = 1.5 \times 10^{-4}\,\mathrm{K}^{-1}$ and $\dot{u}_0 = 0.5\,\mathrm{mms}^{-1}$.

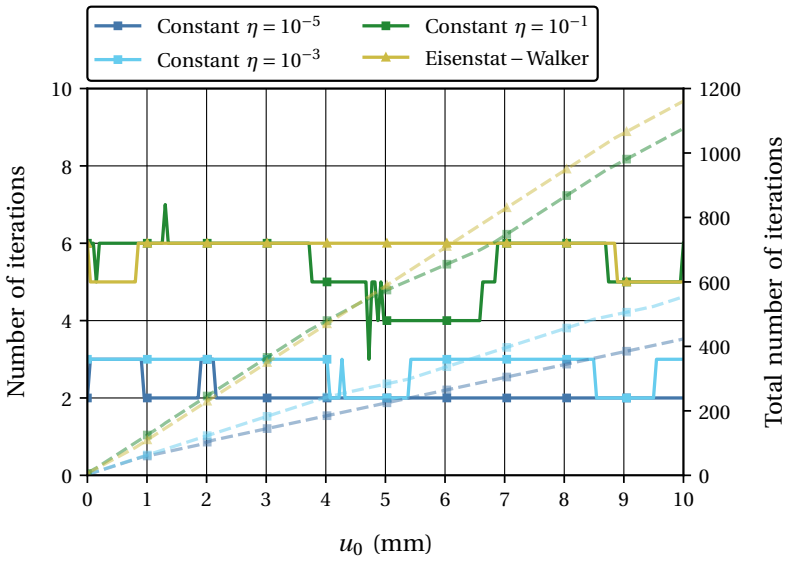


Figure 8.15: Number of nonlinear iterations needed to solve the coupled problem at each time step and the total number of iterations needed to solve the coupled problem for the Newton-GMRES method with a constant forcing term ($\eta = 10^{-5}$, 10^{-3}, 10^{-1}) and the Eisenstat-Walker scheme in the solution of the quasi-static expansion of a thermoelastic thick-walled cylinder with $\alpha_T = 1.5 \times 10^{-4}\,\mathrm{K}^{-1}$ and $\dot{u}_0 = 0.5\,\mathrm{mms}^{-1}$.

a constant forcing term equal to 10^{-1}.

Figures 8.15 presents the total number of residual evaluations a function of the thermal expansion coefficient. The most efficient Newton-GMRES methods analyzed are the ones using a constant forcing term equal to 10^{-5} and 10^{-3}, followed by the technique employing a constant forcing term equal to 10^{-1} and then by the method utilizing the Eisenstat-Walker scheme.

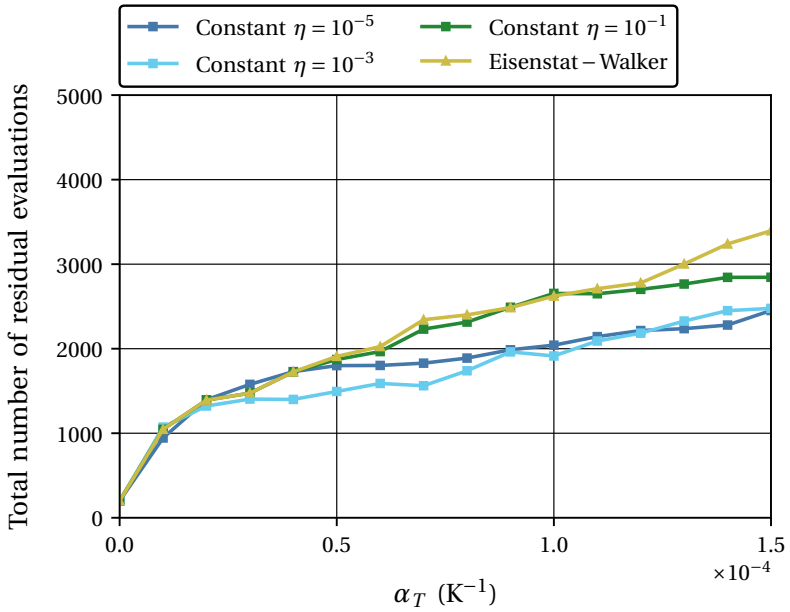


Figure 8.16: Total number of residual evaluations as a function of the thermal expansion coefficient for the Newton-GMRES method with a constant forcing term ($\eta = 10^{-5}$, 10^{-3}, 10^{-1}) and the Eisenstat-Walker scheme in the solution of the quasi-static expansion of a thermoelastic thick-walled cylinder with $\alpha_T = 0\,\mathrm{K}^{-1}$ to $1.5 \times 10^{-4}\,\mathrm{K}^{-1}$ and $\dot{u}_0 = 0.5\,\mathrm{mms}^{-1}$.

8.1.2.4 Polynomial vector extrapolation in cycling mode

The polynomial vector extrapolation methods in cycling mode considered are the MPE and RRE, restricted to at most five evaluations of the residual function per nonlinear iteration. Thus, the combinations analyzed are characterized by the ordered pairs $(n,k) = (1,1)$, $(1,2)$, $(1,3)$, $(2,1)$, $(2,2)$ and $(3,1)$ (see Section 7.5.3).

Figure 8.17 presents the residual in percentage as a function of the number of nonlinear iterations in the first time step with $\alpha_T = 1.5 \times 10^{-4}\,\mathrm{K}^{-1}$ and $\dot{u}_0 = 0.5\,\mathrm{mm\,s}^{-1}$. In general, the larger the number of residual evaluations per nonlinear iteration, the faster rate of convergence. Also, it seems that for the same number of residual evaluations per nonlinear iterations, e.g., $(n,k) = (1,3)$, $(2,2)$ and $(3,1)$ or $(n,k) = (1,2)$, $(2,1)$, it is more profitable to increase k than n. The results for the MPE and the RRE are very similar.

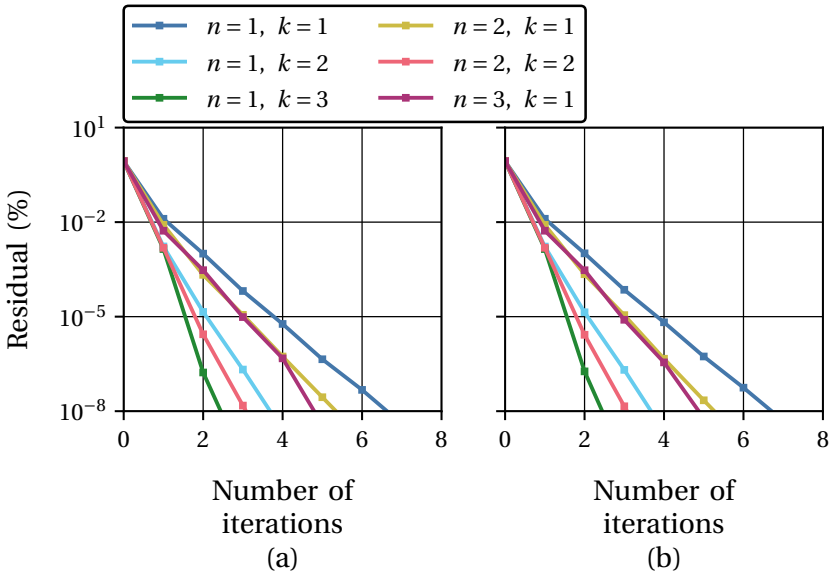


Figure 8.17: Residual in percentage as a function of the number of nonlinear iterations in the first time step for the polynomial vector extrapolation methods in cycling mode, MPE and RRE, restricted to at most five evaluations of the residual function per nonlinear iteration in the solution of the quasi-static expansion of a thermoelastic thick-walled cylinder with $\alpha_T = 1.5 \times 10^{-4}\,\mathrm{K}^{-1}$ and $\dot{u}_0 = 0.5\,\mathrm{mm\,s}^{-1}$.

Figure 8.18 presents the number of residual evaluations needed to solve the coupled problem at each time step and the total (cumulative) number of iterations needed. The number of nonlinear iterations needed to solve the coupled problem to the desired accuracy at each time step remains approximately constant, with a slight decrease as the displacement increases. As before, the MPE and RRE methods display very similar performances. As hinted by the previous results regarding the residual, the methods with more residual evaluations per nonlinear iteration take, in general, fewer iterations to converge.

Figures 8.18 presents the total number of residual evaluations as a function of the thermal expansion coefficient. There is not a clear winner regarding the total number

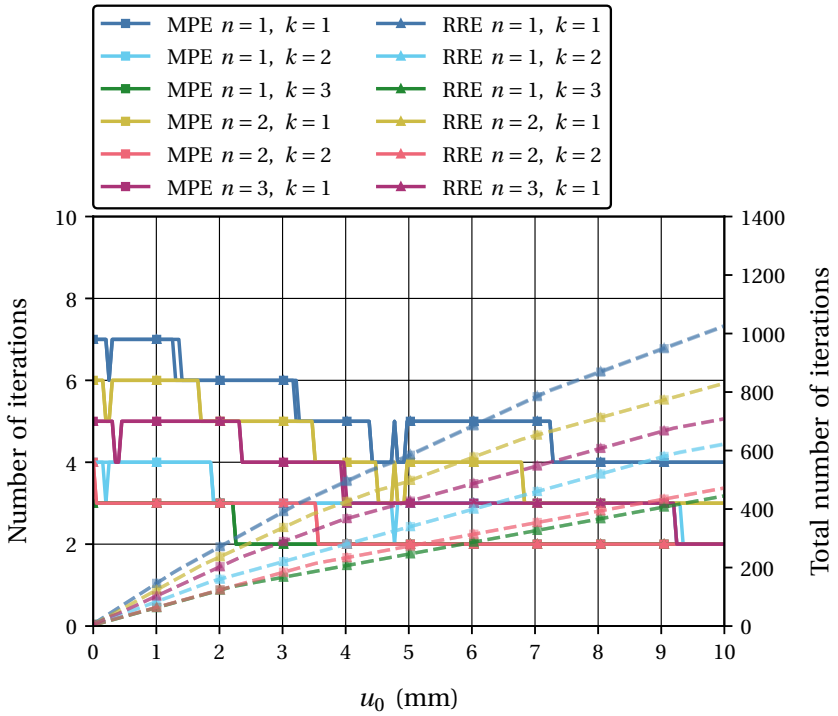


Figure 8.18: number of nonlinear iterations needed to solve the coupled problem at each time step and the total number of iterations needed to solve the coupled problem for the polynomial vector extrapolation methods in cycling mode, MPE and RRE, restricted to at most five evaluations of the residual function per nonlinear iteration in the solution of the quasi-static expansion of a thermoelastic thick-walled cylinder with $\alpha_T = 1.5 \times 10^{-4}\,\mathrm{K}^{-1}$ and $\dot{u}_0 = 0.5\,\mathrm{mm\,s}^{-1}$.

of residual evaluations taken to solve the coupled thermomechanical problem completely, but the methods with $(n,k) = (1,3)$, $(1,2)$ and $(2,2)$ seem to always be in the top spots regarding efficiency.

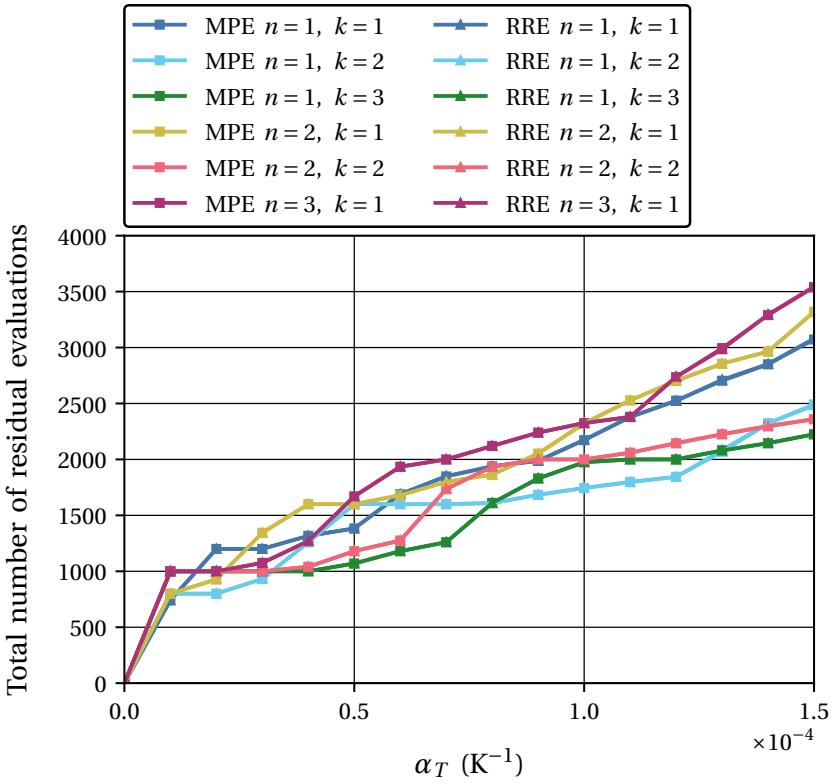


Figure 8.19: Total number of residual evaluations as a function of the thermal expansion coefficient for the polynomial vector extrapolation methods in cycling mode, MPE and RRE, restricted to at most five evaluations of the residual function per nonlinear iteration in the solution of the quasi-static expansion of a thermoelastic thick-walled cylinder with $\alpha_T = 0\,\mathrm{K}^{-1}$ to $1.5\times10^{-4}\,\mathrm{K}^{-1}$ and $\dot{u}_0 = 0.5\,\mathrm{mms}^{-1}$.

8.1.2.5 comparison of the best methods in each class

In this section, the best-performing methods from each of the classes considered are compared with each other. The implicit methods selected are the Aitken relaxation, Broyden's method with Type I update, a Broyden-like method with $\beta = -1$ and $s = 2$, the Newton-GMRES method with $\eta = 10^{-3}$ and the MPE in cycling mode with $(n,k) = (1,3)$.

Figure 8.20 shows the total CPU time in seconds and the total number of residual evaluations as a function of the thermal expansion coefficient for the best performing implicit methods in each class considered. The best performing methods are Broyden's method and the Broyden-like method with $\beta = -1$, followed by the Aitken relaxation, the MPE in cycling mode, and the Newton-GMRES method. As the thermal coefficient, and thus the strength of the coupling, increases, all the methods display a

loss in efficiency. Table 8.2 surmises all the results previously discussed regarding the computational time taken by each solver.

Figure 8.21 depicts the total CPU time in seconds, and time profile, as a function of the mesh size. For all solvers and mesh sizes, evaluating the solvers is the operation that takes up the most computational time. The computation of the residual function considered implies the solution of both the mechanical and thermal problems, with the mechanical solver taking longer than the thermal solver. This difference can be partly explained because it has double the number of degrees of freedom. Also, the operations concerning the constitutive behavior of the material are performed in the mechanical solver for the implementation developed in this work. The time spent on operations concerning the coupling solver solely contributes a small weight to the total time for all mesh sizes and solvers. This relative importance of the operations in the coupling solver is verified despite the volumetric coupling between the mechanical and thermal problems. A volumetric coupling leads to all the degrees of freedom in one of the solvers, in this case, the thermal solver, to have to be considered by the coupling procedure. In contrast, only the degrees of freedom at the contact surface need to be considered for surface coupling, such as the one found in fluid-structure interaction.

Table 8.2: Total CPU time in seconds and total number of residual evaluations as a function of the thermal expansion coefficient for the best performing implicit methods in each class considered in the solution of the quasi-static expansion of a thermoelastic thick-walled cylinder with $\alpha_T = 5\times10^{-5}\,\mathrm{K}^{-1}$, $10\times10^{-5}\,\mathrm{K}^{-1}$, and $15\times10^{-5}\,\mathrm{K}^{-1}$, and $\dot{u}_0 = 0.5\,\mathrm{mm\,s}^{-1}$.

	CPU Time (s)			Nr Residual Evaluations		
α_T $(10^{-5}\,\mathrm{K}^{-1})$	5	10	15	5	10	15
AITK	2.57×10^2	4.03×10^2	5.58×10^2	1025	1378	1796
BRDI	2.20×10^2	3.65×10^2	4.64×10^2	836	1226	1494
BRDI2	2.20×10^2	3.35×10^2	4.65×10^2	836	1108	1466
NEWT	3.11×10^2	4.71×10^2	6.58×10^2	1294	1714	2277
MPE	2.73×10^2	5.59×10^2	6.76×10^2	1070	1975	2225

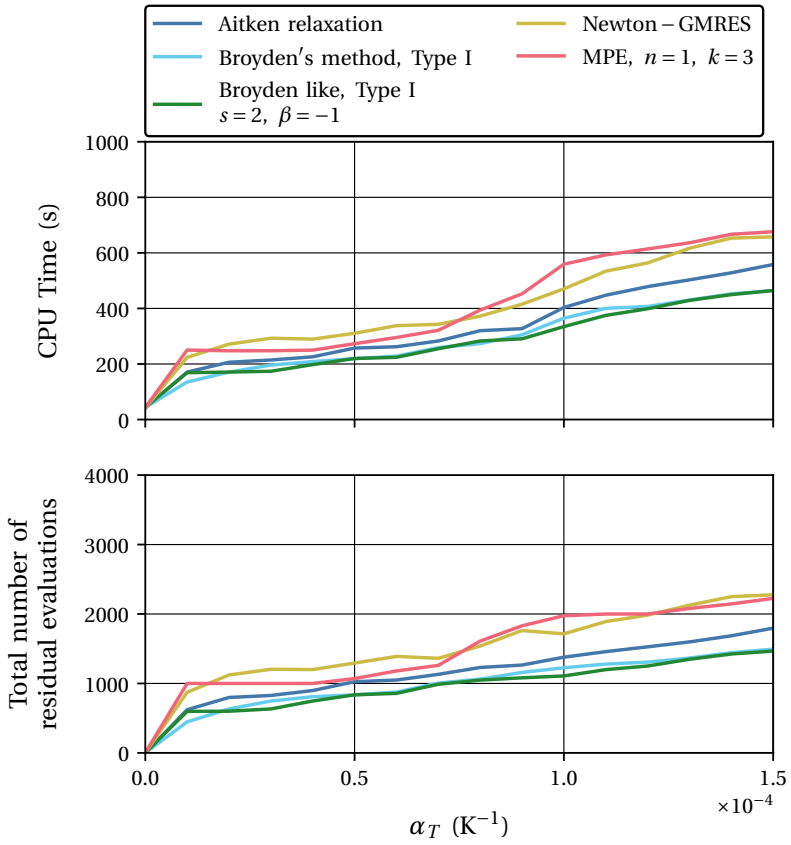


Figure 8.20: Total CPU time in seconds and the total number of residual evaluations as a function of the thermal expansion coefficient for the best performing implicit methods in each class considered in the solution of the quasi-static expansion of a thermoelastic thick-walled cylinder with $\alpha_T = 0\,K^{-1}$ to $1.5\times10^{-4}\,K^{-1}$ and $\dot{u}_0 = 0.5\,mms^{-1}$.

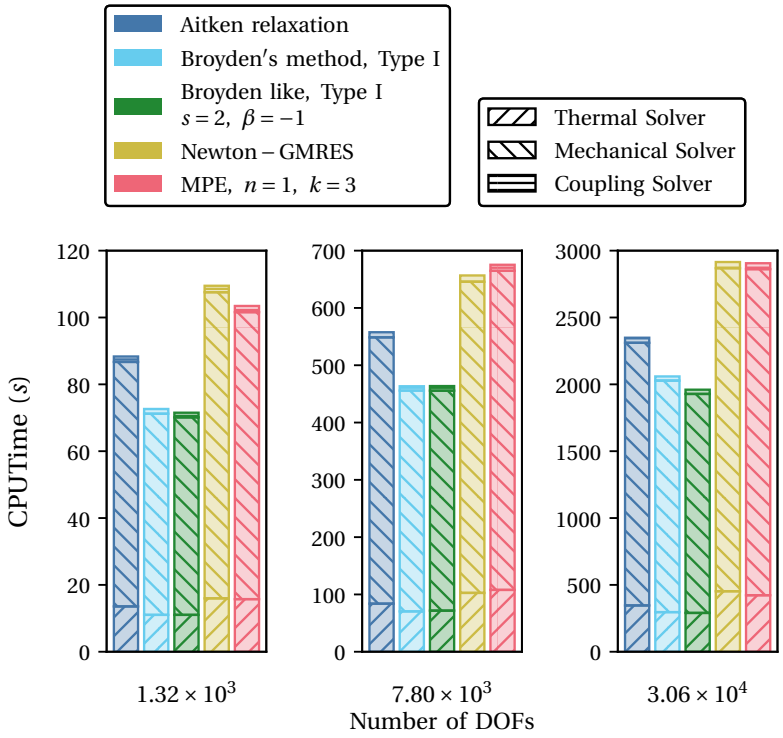


Figure 8.21: Total CPU time in seconds, and time profile, as a function of the mesh size for the best performing implicit methods in each class considered in the solution of the quasi-static expansion of a thermoelastic thick-walled cylinder with $\alpha_T = 1.5 \times 10^{-4}\,\mathrm{K}^{-1}$ and $\dot{u}_0 = 0.5\,\mathrm{mm\,s}^{-1}$.

8.1.2.6 Effect of predictiors

Figure 8.22 presents the effect of employing a linear or quadratic predictor on the number of residual evaluations needed to fully solve the thermomechanical problem under analysis as a function of the thermal expansion coefficient. All methods improve using the polynomial predictors, with the best results achieved using the quadratic predictor. The decrease in the number of residual evaluations is around 30% to 40%. Their effect seems to display a slight increase as the coupling gets stronger.

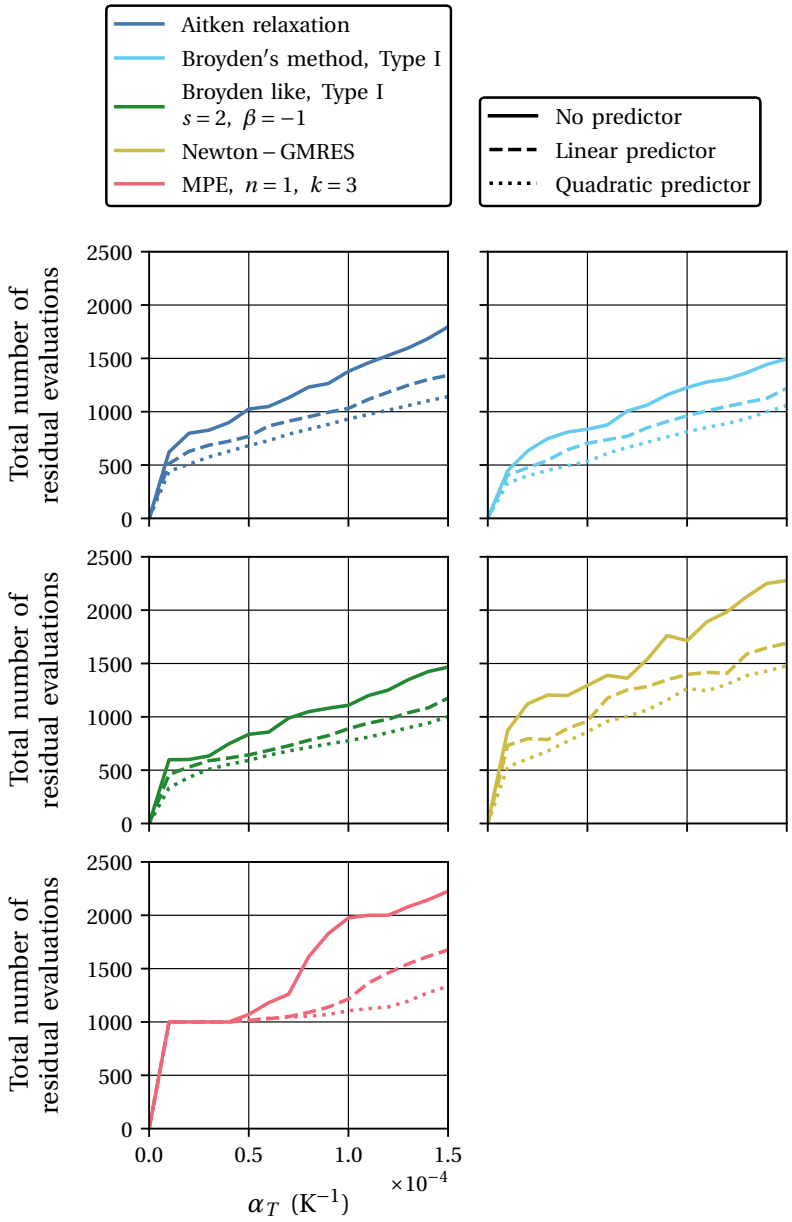


Figure 8.22: Total number of iterations as a function of the thermal expansion coefficient for the best performing implicit methods in each class considered using a linear, a quadratic, and no predictor in the solution of the necking of a circular bar with $\alpha_T = 0\,\mathrm{K}^{-1}$ to $1.5 \times 10^{-4}\,\mathrm{K}^{-1}$ and $\dot{u}_0 = 0.5\,\mathrm{mm\,s}^{-1}$.

8.2 Necking of a circular bar

The second validation example consists of the thermally triggered necking of a circular bar, as initially reported in **?** and replicated in **?**. The problem consists of a cylindrical bar of radius $r = 6.413\,\text{mm}$ and length $h = 53.334\,\text{mm}$ subject to a prescribed axial displacement $\bar{u}_y = 8\,\text{mm}$ at both ends during $t = 8\,\text{s}$. The supports at the tips allow transverse contraction of the specimen. The bar is initially at the ambient temperature $T_0 = T_\infty = 293\,\text{K}$ and is subject to heat transfer by convection at all boundaries, with a heat transfer coefficient $h_c = 17.5 \times 10^{-3}\,\text{N}\,\text{mm}^{-1}\,\text{s}^{-1}\,\text{K}^{-1}$. The material is modeled with the constitutive model by **?** and the material properties are given in Table 8.3.

This classical benchmark in isothermal elastoplasticity renders a bifurcation problem where a geometric imperfection typically triggers the necking phenomenon. In the thermomechanical version, the combination of plastic dissipation in the bulk and heat transfer at the boundaries produce a temperature field that becomes progressively more heterogeneous during the loading. With growing elongation, the temperature rise in the center of the bar will increase relative to the exterior boundary and automatically trigger the necking, even for a geometrically perfect setup.

The problem is analyzed using two-dimensional axisymmetric QUAD4 elements for both the mechanical and the thermal problems. Figure 8.23 illustrates the problem setup, the finite element mesh employed in the 2D simulations, and distinct stages of deformation and temperature field during the prescribed elongation, evidencing the significant necking of the bar. Only one-quarter of the specimen is simulated, resulting in a finite element mesh with 1326 nodes and 1250 elements. Except where explicitly indicated, this is the mesh employed. The load is applied in 80 equal time steps $\Delta t = 0.1\,\text{s}$. Using a backward Euler integration, a quasi-static solution is computed for the mechanical problem. The transient temperature field is integrated with the generalised-α method with $\rho_{\infty,T} = 1.0$.

8.2.1 Validation of the Numerical Results

As validation for the results presented in the present work, the reaction force at the supports and the neck surface temperature at point A are compared to results found in the literature, see 8.23. Different reference data is included in the analysis, particularly the adiabatic and non-adiabatic results presented in the original work by **?**, and the non-adiabatic results presented in **?**. It should be remarked that there are five fundamental differences between these two publications that lead to distinct results. First, in **?**, the authors adopt a mechanical dissipation term that is thermodynamically inconsistent, based on the previously mentioned dissipation factor, χ, whereas, in **?**, the authors use the mechanical dissipation coming directly from the second law of thermodynamics. Second, **?** also uses a consistent structural heating term for the Gough-Joule effect, accounting for both elastic and plastic contributions. In contrast, **?** considered the elastic contributions, exclusively. For more information on the previous classification and mathematical formulas for the heating parcels, the reader is referred to **??**. The third difference is linked to the heat conduction law employed in each contribution. Although the large deformation version of the Fourier law underlies both publications, **?** considered as a fixed material parameter the spatial thermal conductivity, k, and in **?**, the material thermal conductivity, k_0, was instead interpreted as the fixed material parameter. Without further mention, the latter is employed in the current work. Fourth, **?** solved the coupled problem using an operator split scheme and **?** pursued a monolithic solution.

Table 8.3: Material properties and initial and boundary conditions for the problem concerning the quasi-static finite strain thermo-elastic expansion of an infinitely long thick-walled cylinder.

Material Properties			Effective value
Density	ρ	$(\mathrm{Ns^2mm^{-4}})$	7.8×10^{-9}
Bulk modulus	κ	$(\mathrm{Nmm^{-2}})$	164.206×10^{3}
Shear modulus	μ	$(\mathrm{Nmm^{-2}})$	801.938×10^{3}
Conductivity	k	$(\mathrm{Ns^{-1}K^{-1}})$	45
Heat capacity	C_V	$(\mathrm{mm^2s^{-2}K^{-1}})$	460×10^{6}
Coefficient of thermal expansion	α_T	$(\mathrm{K^{-1}})$	10×10^{-6}
Dissipation factor	χ	(-)	900×10^{-3}
Initial yield stress at T_0	$\sigma_{y,0}$	$(\mathrm{Nmm^{-2}})$	450
Linear hardening coefficient at T_0	H	$(\mathrm{Nmm^{-2}})$	129.24
Saturation exponent	δ	(-)	16.93
Saturation yield stress at T_0	$\sigma_{y,\infty}$	$(\mathrm{Nmm^{-2}})$	715
Thermal softening parameter ($\sigma_{y,0}$)	ω_0	$(\mathrm{K^{-1}})$	2×10^{-3}
Thermal softening parameter ($\sigma_{u,\infty}, H$)	ω_h	$(\mathrm{K^{-1}})$	2×10^{-3}
Boundary Conditions			
Radius of the cylindrical bar	r	(mm)	6.413
Length of the cylindrical bar	h	(mm)	53.334
Maximum displacement at both ends	$\bar{u}_y$	(mm)	8
Time to maximum displacement	t	(s)	8
Heat transfer coefficient	h_c	$(\mathrm{Nmm^{-1}K^{-1}})$	17.5×10^{-3}
Initial Conditions			
Initial temperature	T_0	(K)	293
Reference value			
Temperature at outer radius ($r = r_0$)	(K)		

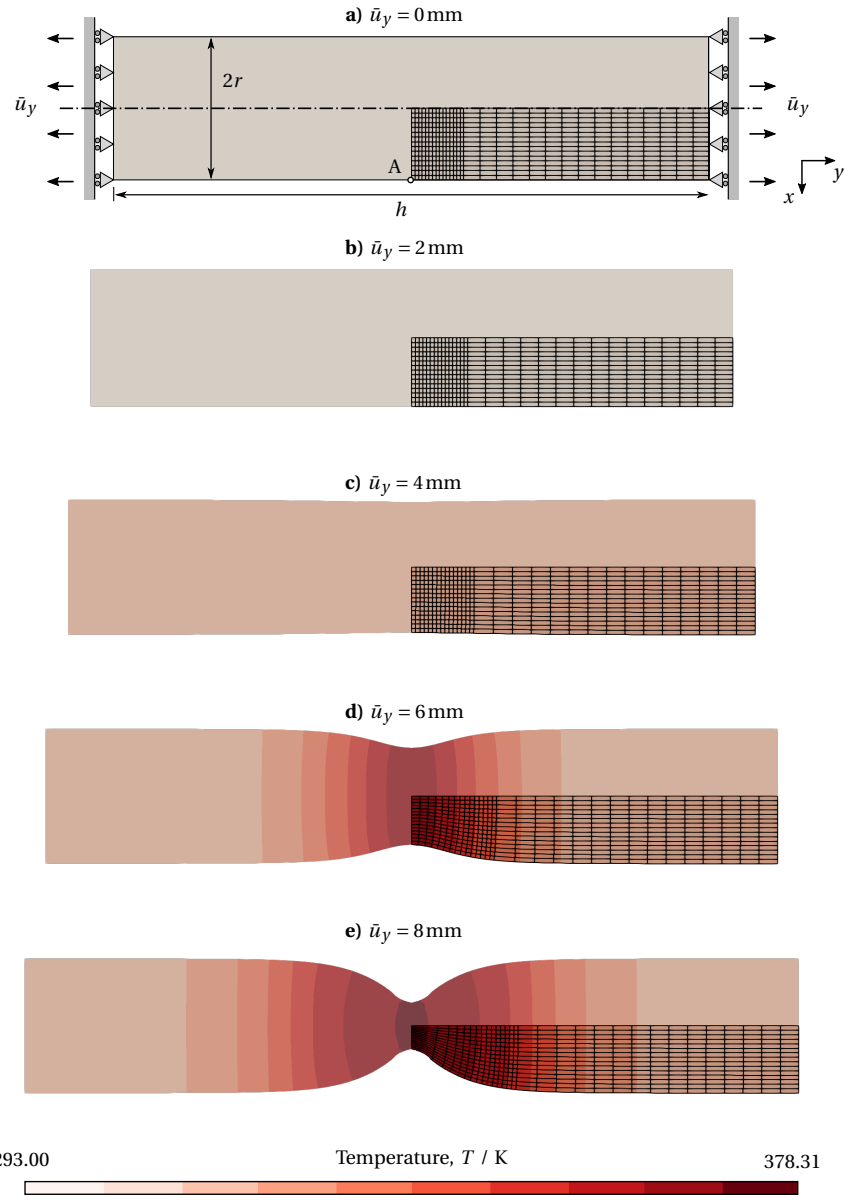


Figure 8.23: Description of the thermally triggered necking of a circular bar problem, characteristic deformation, and temperature field stages during the loading and example axisymmetric finite element mesh. The results have been obtained with QUAD4 elements, non-adiabatic boundary conditions, the inconsistent mechanical dissipation formulation, and the Fourier law based on constant k_0.

Last, regarding spatial discretisation, **?** used HEXA8-FBAR elements and **?** employed mixed displacement-pressure QUAD8 elements.

To enable a fair comparison with the reference data, the numerical solution is calculated using both the adiabatic and non-adiabatic setups. Furthermore, consistent and inconsistent interpretations are considered in the latter, including a calculation with fixed k. The evolution of the reaction force and neck surface temperature as a function of the prescribed displacement are shown in 8.24 for the 2D and 3D solutions, respectively. From a physical interpretation standpoint, the plot of reaction forces suggests that the simulation occurs almost entirely in the elastoplastic regime. The necking process does not occur in the isothermal and adiabatic solutions, with the adiabatic solution predicting slightly smaller reactions due to the thermal softening effect. Also, in this case, the temperature evolves uniformly in the bar and grows nearly linearly over time, as captured in the numerical solutions. As previously postulated, the necking is automatically triggered in the non-adiabatic solutions, which produces a heavy reduction in the reaction force starting approximately at $\bar{u}_y = 4\,\text{mm}$, followed by a steep temperature rise due to the higher plastic dissipation.

Inspecting 8.24 from a validation perspective, the numerical results show a good agreement with the literature. The correlation between the non-adiabatic, inconsistent solution and fixed k with the results from **?** is very satisfactory, both on the temperature and the reaction force side. Curiously, while the reaction force seems almost insensitive to the element type employed, the neck temperature is better approximated with the QUAD8 solution, especially near the maximum displacement. The slight difference between these cases can presumably be attributed to the distinct element technology and coupling solution strategy used to obtain the two curves. Relative to the solutions based on the consistent mechanical heating, the numerical results show good agreement with the curves extracted from **?** up to $\bar{u}_y = 3\,\text{mm}$, but features significant differences from there on, both in the mechanical and thermal responses. The results are consistent insofar as the numerical solution obtained predicts a smaller temperature increase but larger reaction forces, as the thermal softening is less pronounced. Unfortunately, to the author's knowledge, no other bibliographical sources consider the entirely consistent thermomechanical version of the model to support any of the sides. Nevertheless, as the accuracy relative to the results from **?** is already adequate, the possibly remaining issue resides at the constitutive model level and therefore does not compromise the coupling environment. It should also be remarked that if the prescribed displacement is slightly larger, the weakly coupled partitioned solution diverges at some point. This can be expected from the mathematical properties of this type of strategy, as discussed in this chapter. In truth, numerical divergence can be observed to start near $\bar{u}_y = 8\,\text{mm}$ for the non-adiabatic, inconsistent solution, see 8.24. In principle, it is possible to stabilize the solution by employing more advanced techniques, for instance, implicit coupling strategies with numerical acceleration. Despite being paramount for a robust computer simulation tool, these topics are postponed to future developments. Overall, the present results are a sound indication of the correct implementation of the coupling environment, particularly the data exchange and solution orchestration.

Figure 8.24: evolution of the reaction force at the tips of the bar and the neck surface temperature with the prescribed displacement using QUAD8 elements with reduced integration (QUAD8R) and HEXA8-FBAR elements.

8.2.2 Evaluation and comparison of implicit solution methods for the coupled problem

The following contains the results concerning the evaluation and comparison of the implicit solution methods for coupled problems considered in this work. The discussion starts with the methods that require only one evaluation of the residual, implicitly requiring the solution of both the thermal and the mechanical problems by the corresponding solvers per nonlinear iteration. The Broyden-like family of methods fall into this category too but are considered by themselves and are presented next. The Newton-Krylov methods are also analyzed, followed by the vector extrapolation methods in cycling mode. The discussion ends with a comparison of the best methods in each class.

As for the thermoelastic case, the analysis presented is based mainly on three pieces of information. The first concerns how the residual evolves as a function of the nonlinear iterations. The second is the number of function evaluations needed to solve the coupled problem to the desired accuracy at each time step. The third is the total number of residual evaluations needed to solve the coupled problem to the desired accuracy as a function of the thermal softening parameters, w_0 and w_h, set to be equal. The larger the softening parameters w_0 and w_h, the stronger the coupling between the thermal and mechanical fields, and the more challenging the problem is to solve.

The maximum displacement at both ends is restricted to 5 mm applied in 5 s to prevent an excessive elongation of the elements in finite element simulation that often leads to simulation failure. The thermal softening parameters $w_0 = w_h$ vary from $2 \times 10^{-3}\,\mathrm{K}^{-1}$ to $10^{-2}\,\mathrm{K}^{-1}$, such that the larger the thermal softening parameters w_0 and w_h, the stronger the coupling. This can be understood from Figure 8.25. In this case, it shows the total number of residual evaluations corresponding to the total number of nonlinear iterations as a function of the thermal softening for the fixed-point method. As w_0 increases, the displacement at which the coupling is the most substantial moves to the left, and the width of the corresponding peak gets larger, employing a stronger coupling.

8.2.2.1 Methods with only one residual evaluation per iteration

The methods with only one residual evaluation per nonlinear iteration considered are the fixed-point method (see Section 7.3.1), the underrelaxation method (see Section 7.3.3), the Aitken relaxation (see Section 7.4.1) and Broyden's method, Type I and II (see Section 7.4.2). These are, a priori, the most economical methods regarding residual evaluations, as only one is performed per nonlinear iteration. The underrelaxation is performed with $\omega = 0.5$, and the first relaxation coefficient for the Aitken relaxation is also set to 0.5.

Figure 8.26 presents the number of nonlinear iterations/residual evaluations needed to solve the coupled thermomechanical problem at each time step and the total (cumulative) number of iterations needed with $w_0 = w_h = 10^{-2}\,\mathrm{K}^{-1}$. It is clear from the number of nonlinear iterations needed to solve the problem at $u_y \approx 2.3\,\mathrm{mm}$ that the coupling is the strongest at this moment. It corresponds to the moment where the necking of the bar begins. There is a marked performance difference between the methods under analysis, with the Broyden methods taking much fewer iterations, 10, than either the Aitken relaxation, 18, or the fixed-point method, 34. For all other moments, the performance of the implicit methods under analysis is very

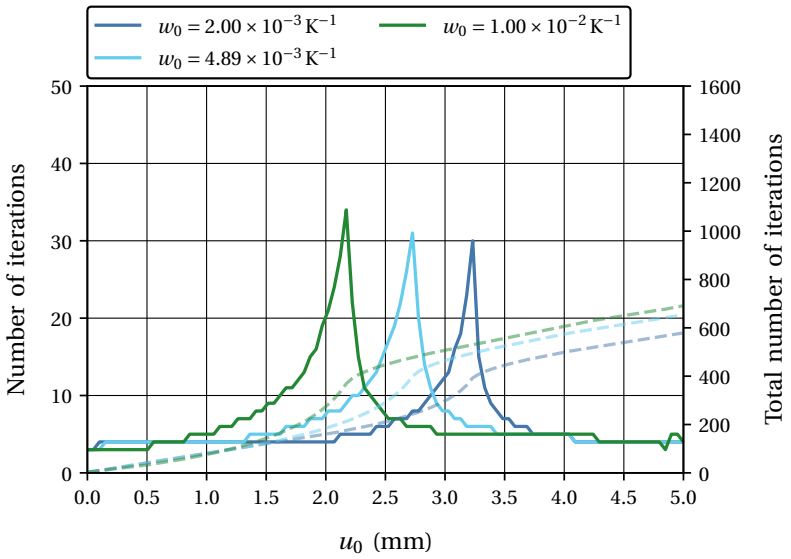


Figure 8.25: Total number of residual evaluations as a function of the thermal expansion coefficient for the fixed-point method in the solution of the necking of a circular bar with $w_0 = w_h = 2 \times 10^{-3}\,\mathrm{K}^{-1}$, $4.89 \times 10^{-3}\,\mathrm{K}^{-1}$, and $10^{-2}\,\mathrm{K}^{-1}$.

similar, except for the under relaxation. This behavior is probably due to the poor choice of the relaxation coefficient.

Figure 8.27 presents the total number of residual evaluations, in this case, corresponding also to the total number of nonlinear iterations, as a function of the thermal softening parameters, which control the strength of the coupling between the thermal and the mechanical fields, respectively. The most efficient methods are the two Broyden methods, followed by the Aitken relaxation. The difference between the total number of iterations is around 100 for the three different methods.

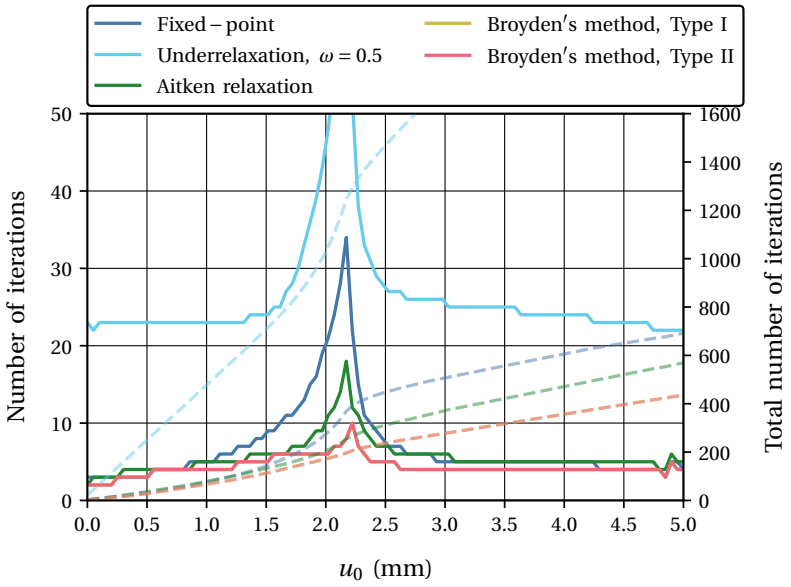


Figure 8.26: number of nonlinear iterations needed to solve the coupled problem at each time step and the total number of iterations needed to solve the coupled problem for the implicit methods that perform only one evaluation per nonlinear iteration in the solution of the necking of a circular bar with $w_0 = w_h = 10^{-2}\,\mathrm{K}^{-1}$.

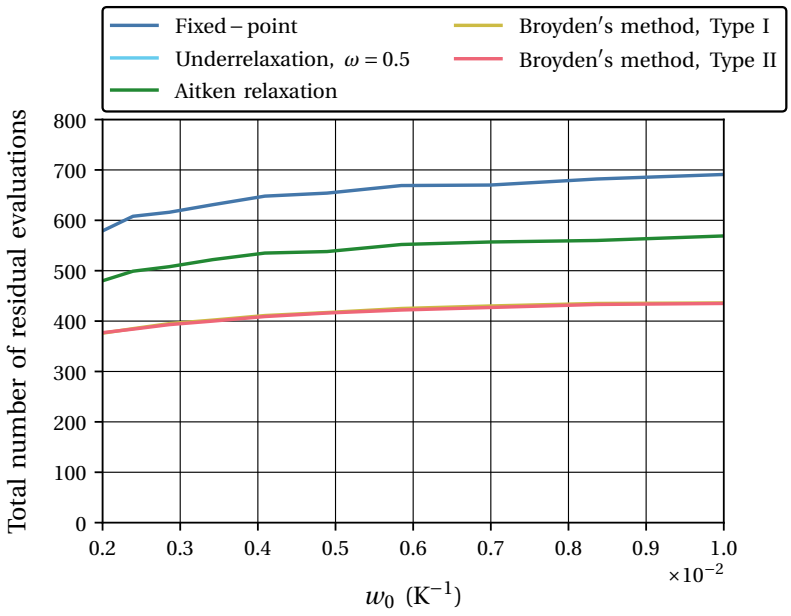


Figure 8.27: Total number of residual evaluations as a function of the thermal expansion coefficient for the implicit methods that perform only one evaluation per nonlinear iteration in the solution of the necking of a circular bar with $w_0 = w_h = 2 \times 10^{-3}\,\mathrm{K}^{-1}$ to $10^{-2}\,\mathrm{K}^{-1}$.

8.2.2.2 Broyden-like method

The Broyden-like methods considered (see Section 7.4.2) employ as group sizes s = 1, 2, 3, 6 with a maximum number of previous iterations available equal to 6. The mixing parameters considered are $\beta = -1$, 2×10^{-3}, 2×10^{-2}. All combinations are also considered with both Type I and Type II updating for the approximation to the Jacobian.

Figures 8.28 and 8.29 present the number of nonlinear iterations/number of function evaluations needed to solve the coupled problem at each time step and the total (cumulative) number of iterations needed in the solution of the necking of a circular bar with $w_0 = w_h = 10^{-2}\,\mathrm{K}^{-1}$. Regarding the methods using $\beta = -1$, the difference between Type I and II methods is not noticeable. However, the different size groups lead to different behaviors. A larger group size leads to worse results for the displacement values where the coupling is the weakest. On the other hand, the number of iterations where the coupling is the strongest are about the same for all group sizes. For $\beta = 2 \times 10^{-3}$ and 2×10^{-2} the only group size that converges is $s = 6$. For the other group sizes, the simulation breaks down approximately after reaching $u_y = 2.3\,\mathrm{mm}$. This failure happens because the mechanical solver fails with an exploding residual in its internal Newton-Raphson procedure. It doesn't seem to be directly related to the coupling solver itself.

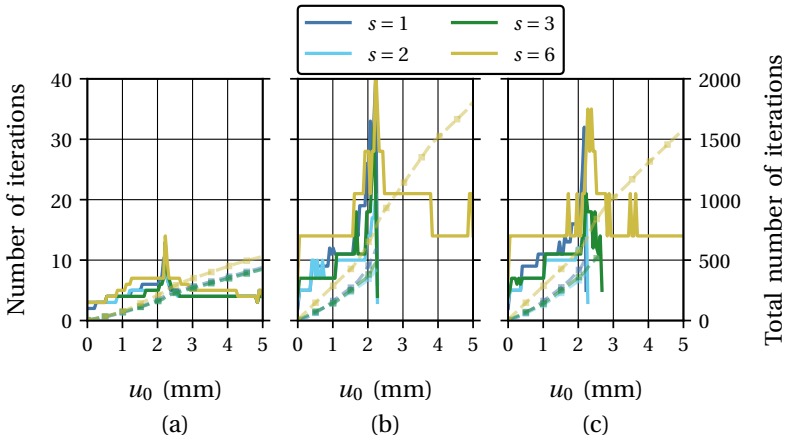


Figure 8.28: Number of nonlinear iterations needed to solve the coupled problem at each time step and the total number of iterations needed to solve the coupled problem for Broyden-like methods with Type I update and group sizes s = 1, 2, 4 and 6: (a) $\beta = -1$, (b) $\beta = 2 \times 10^{-3}$, and (c) $\beta = 2 \times 10^{-2}$ in the solution of the necking of a circular bar with $w_0 = w_h = 10^{-2}\,\mathrm{K}^{-1}$.

Figures 8.30 presents the total number of residual evaluations as a function of the thermal softening parameter, respectively, for the Broyden-like methods with Type I update considered. The same results are presented in Figures 8.31 for the Broyden-like methods with Type II update. For $\beta = -1$, the best results are obtained for s = 1, 2, and 3. However, convergence is not always achieved for $s = 3$. For $\beta = 2 \times 10^{-3}$ and 2×10^{-3} the best results seem to be obtained for $s = 2$ when the methods do converge. However, this is a rare occurrence.

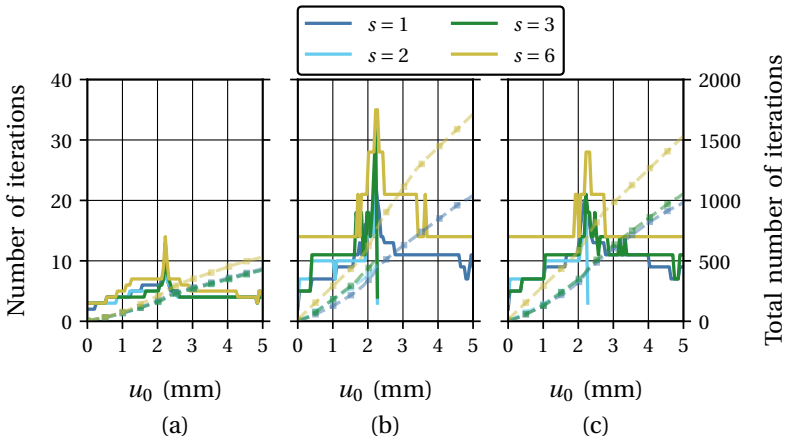


Figure 8.29: Number of nonlinear iterations needed to solve the coupled problem at each time step and the total number of iterations needed to solve the coupled problem for Broyden-like methods with Type II update and group sizes $s = 1$, 2, 4 and 6: (a) $\beta = -1$, (b) $\beta = 2 \times 10^{-3}$, and (c) $\beta = 2 \times 10^{-2}$ in the solution of the necking of a circular bar with $w_0 = w_h = 10^{-2}\,\mathrm{K}^{-1}$.

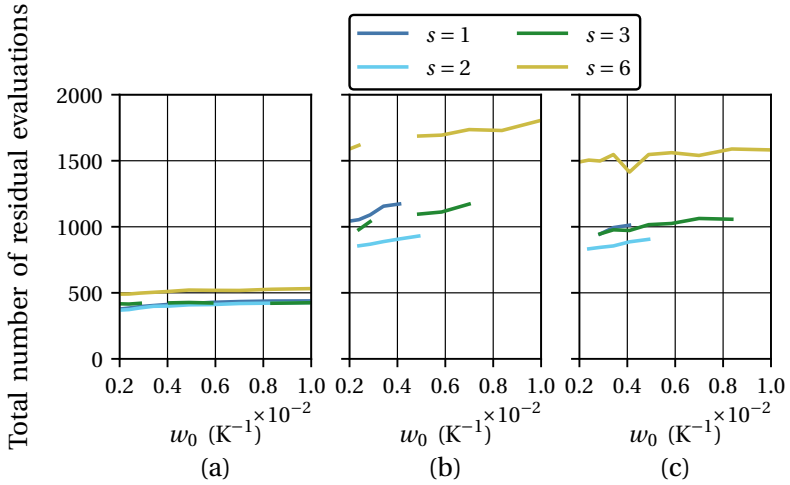


Figure 8.30: Total number of residual evaluations as a function of the thermal expansion coefficient for the implicit methods for Broyden-like methods with Type I update and group sizes $s = 1$, 2, 4 and 6: (a) $\beta = -1$, (b) $\beta = 2 \times 10^{-3}$, and (c) $\beta = 2 \times 10^{-2}$ in the solution of the necking of a circular bar with $w_0 = w_h = 2 \times 10^{-3}\,\mathrm{K}^{-1}$ to $10^{-2}\,\mathrm{K}^{-1}$.

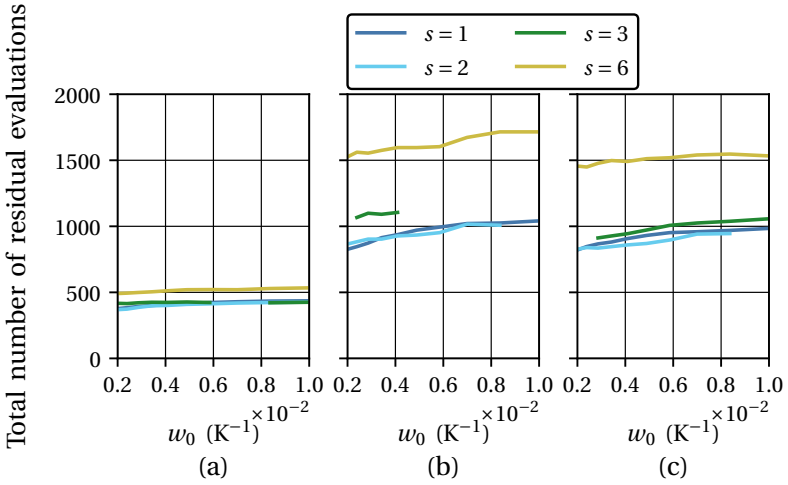


Figure 8.31: Total number of residual evaluations as a function of the thermal expansion coefficient for Broyden-like methods with Type II update and group sizes $s = 1$, 2, 4 and 6: (a) $\beta = -1$, (b) $\beta = 2 \times 10^{-3}$, and (c) $\beta = 2 \times 10^{-2}$ in the solution of the necking of a circular bar with $w_0 = w_h = 2 \times 10^{-3}\,\mathrm{K}^{-1}$ to $10^{-2}\,\mathrm{K}^{-1}$.

8.2.2.3 Newton-GMRES method

The Newton-Krylov method examined here uses as the Krylov subspace solver the GMRES method (see Section 7.5.2). Different forcing terms are employed. The values considered are $\eta = 10^{-1}$, 10^{-3} and 10^{-5}. The Eisenstat-Walker scheme for the adaptive choice of the forcing term is also utilized.

Figure 8.32 presents the number of nonlinear iterations needed to solve the coupled problem at each time step and the total (cumulative) number of iterations needed. For the Newton-GMRES methods considered, the moment where the coupling is the strongest is not as clear as for the methods considered until now. However, the number of iterations needed to solve the thermomechanical problem at each timestep is the largest for $u_y = 2.3\,\mathrm{mm}$. Also, contrary to the other methods, where the coupling is the weakest, there is still some difference between the methods considered. In general, higher forcing terms lead to fewer iterations. However, remember that less iteration doesn't necessarily correspond to fewer residual evaluations.

Figure 8.33 present the total number of residual evaluations as a function of the thermal softening parameters w_0 and w_h. The best performing method employs a constant forcing term equal to 10^{-3}. The other constant forcing terms and the Eisenstat-Walker method display a similar efficiency. The methods using a constant forcing also fail to converge for $w_0 = w_h = 2 \times 10^{-3}\,\mathrm{K}^{-1}$.

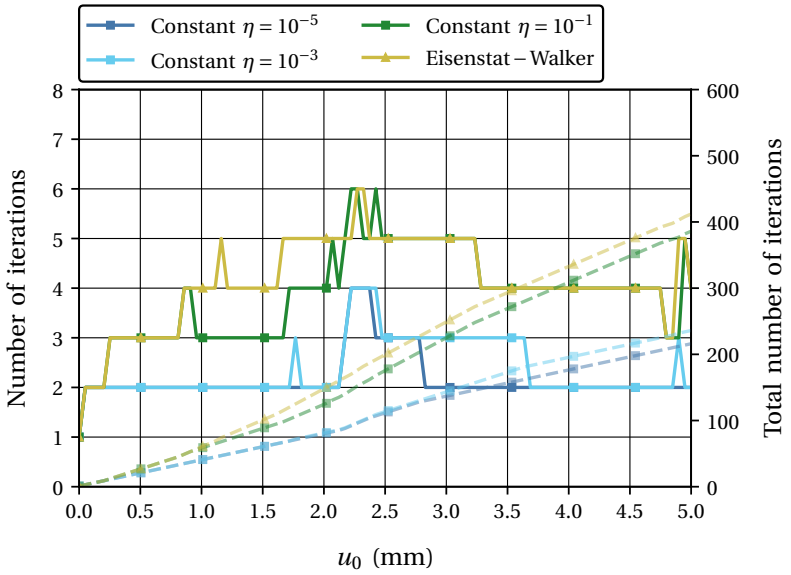


Figure 8.32: Number of nonlinear iterations needed to solve the coupled problem at each time step and the total number of iterations needed to solve the coupled problem for the Newton-GMRES method with a constant forcing term ($\eta = 10^{-5}$, 10^{-3}, 10^{-1}) and the Eisenstat-Walker scheme in the solution of the necking of a circular bar with $w_0 = w_h = 10^{-2}\,\mathrm{K}^{-1}$.

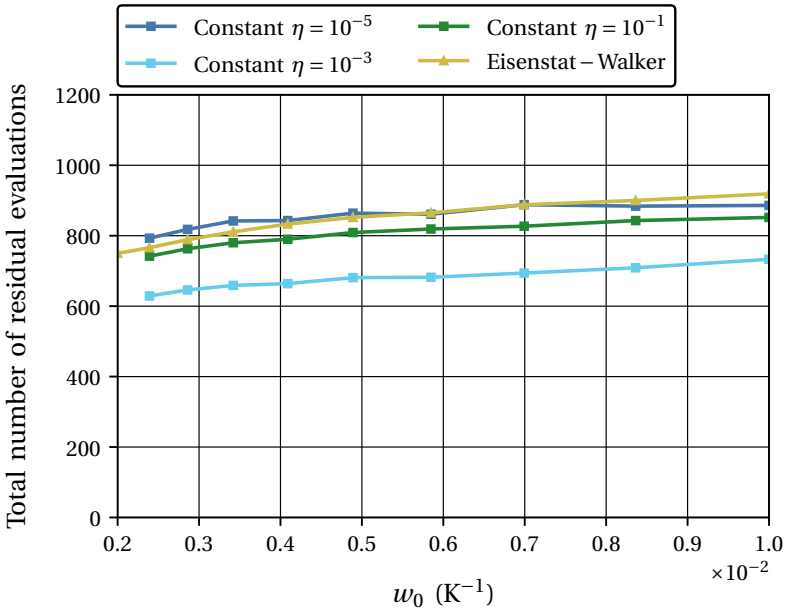


Figure 8.33: Total number of residual evaluations as a function of the thermal expansion coefficient for the Newton-GMRES method with a constant forcing term ($\eta = 10^{-5}$, 10^{-3}, 10^{-1}) and the Eisenstat-Walker scheme in the solution of the necking of a circular bar with $w_0 = w_h = 2 \times 10^{-3}\,\mathrm{K}^{-1}$ to $10^{-2}\,\mathrm{K}^{-1}$.

8.2.2.4 Polynomial vector extrapolation in cycling mode

The polynomial vector extrapolation methods in cycling mode considered are the MPE and RRE, restricted to at most five evaluations of the residual function per nonlinear iteration. Thus, the combinations analyzed are characterized by the ordered pairs $(n,k) = (1,1)$, $(1,2)$, $(1,3)$, $(2,1)$, $(2,2)$ and $(3,1)$.

Figure 8.34 presents the number of residual evaluations needed to solve the coupled problem at each time step and the total (cumulative) number of iterations needed with $w_0 = w_h = 10^{-2}\,\mathrm{K}^{-1}$. There is no clear difference between the MPE and the RRE. The moment of strongest coupling is visible for the methods using few residual evaluations per nonlinear iteration, such as $(n,k) = (1,1)$. For $(n,,k) = (1,3)$, this is much less noticeable. Regarding the moments where the coupling is weaker, the difference between the methods is less marked.

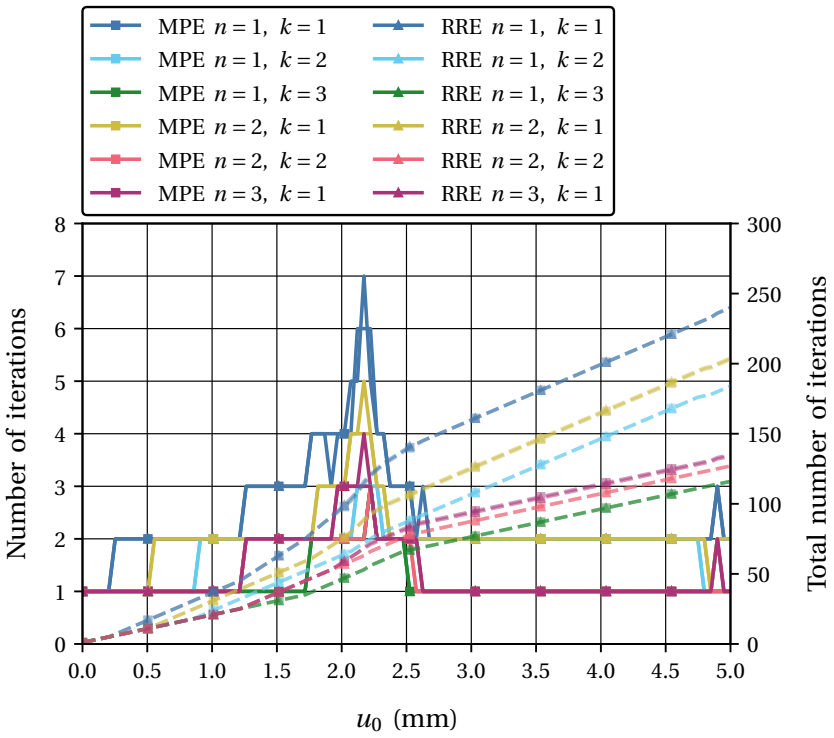


Figure 8.34: number of nonlinear iterations needed to solve the coupled problem at each time step and the total number of iterations needed to solve the coupled problem for the polynomial vector extrapolation methods in cycling mode, MPE and RRE, restricted to at most five evaluations of the residual function per nonlinear iteration in the solution of the necking of a circular bar with $w_0 = w_h = 10^{-2}\,\mathrm{K}^{-1}$.

Figures 8.35 present the total number of residual evaluations as a function of the thermal expansion coefficient. The method which performs the best uses $(n,k) = (1,3)$. The second and third best methods are also the ones using five residual evaluations per nonlinear iteration, $(n,k) = (2,2)$ and $(n,k) = (3,1)$.

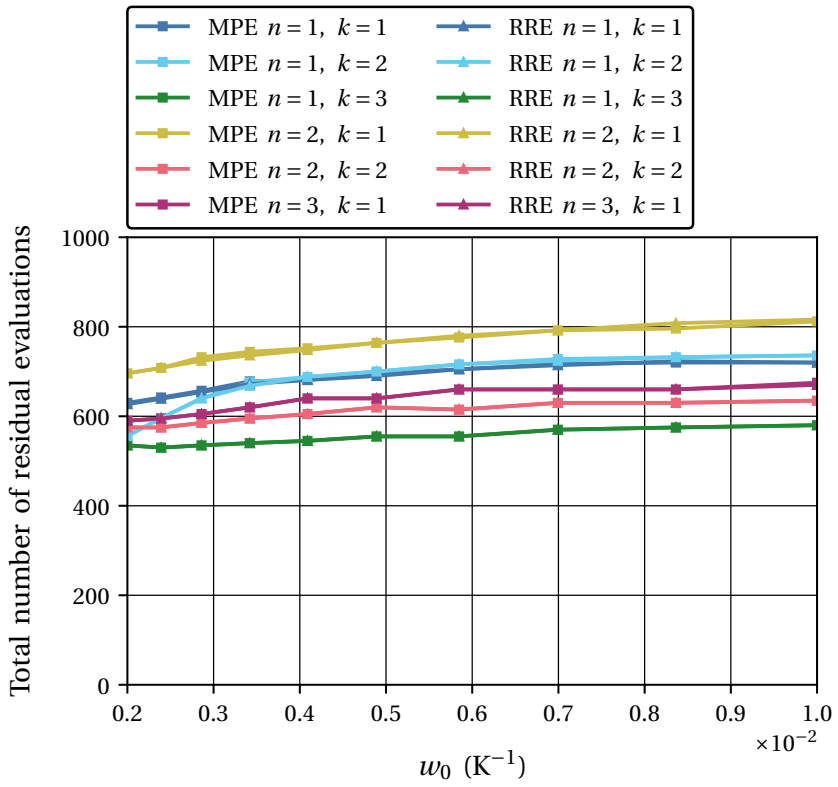


Figure 8.35: Total number of residual evaluations as a function of the thermal expansion coefficient for the polynomial vector extrapolation methods in cycling mode, MPE and RRE, restricted to at most five evaluations of the residual function per nonlinear iteration in the solution of the necking of a circular bar with $w_0 = w_h = 2 \times 10^{-3}\,\mathrm{K}^{-1}$ to $10^{-2}\,\mathrm{K}^{-1}$.

8.2.2.5 comparison of the best methods in each class

In this section, the best-performing methods from each of the classes considered are compared with each other. The implicit methods selected are Aitken relaxation, Broyden's method with Type I update, a Broyden-like method with $\beta = -1$ and $s = 2$, the Newton-GMRES method with $\eta = 10^{-3}$ and the MPE in cycling mode with $(n,k) = (1,3)$.

Figure 8.36 shows the total CPU time in seconds and the total number of residual evaluations as a function of the thermal softening parameters for the best performing implicit methods in each class considered. The best performing methods are Broyden's and the Broyden-like methods, whose performance is very similar. They are followed by the Aitken relaxation and the MPE in cycling mode, also displaying a comparable efficiency. The Newton-GMRES method is the worse performing method considered. As the thermal softening parameters, and thus the strength of the coupling, increase, all the methods display approximately a similar loss in efficiency. Table 8.4 surmises all the results previously discussed regarding the computational time each solvers takes.

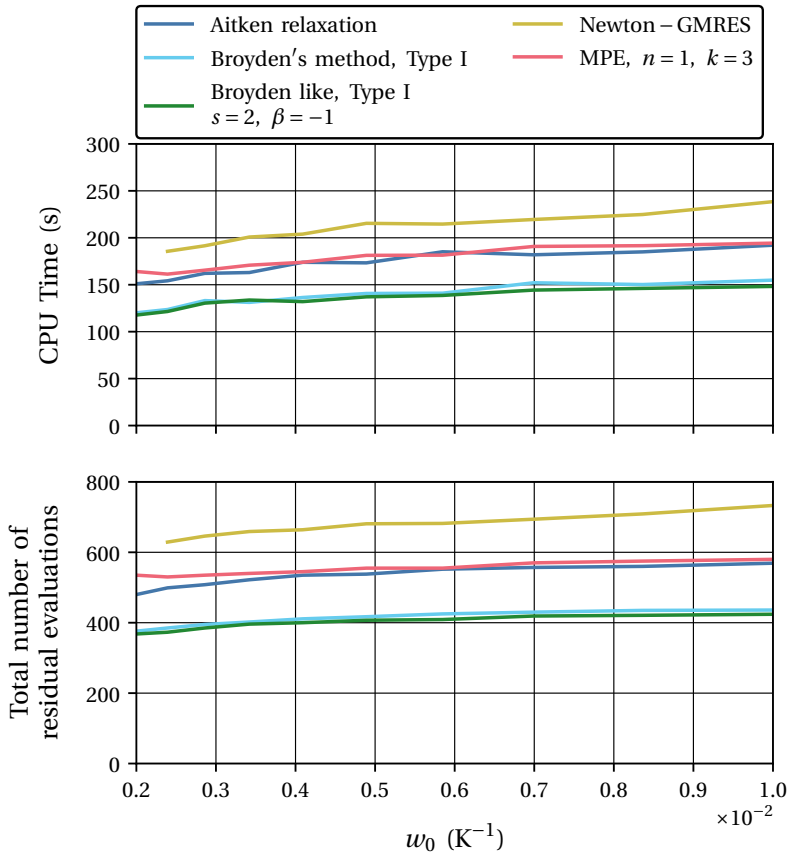


Figure 8.36: Total CPU time in seconds and the total number of residual evaluations as a function of the thermal softening parameters $w_0 = w_h$ for the best performing implicit methods in each class considered in the solution of the necking of a circular bar with $w_0 = w_h = 2 \times 10^{-3}\,\mathrm{K}^{-1}$ to $10^{-2}\,\mathrm{K}^{-1}$.

Table 8.4: Total CPU time in seconds and the total number of residual evaluations as a function of the thermal softening parameters $w_0 = w_h$ for the best performing implicit methods in each class considered in the solution of the necking of a circular bar with $w_0 = w_h = 2 \times 10^{-3}\,\mathrm{K}^{-1}$, $4.89 \times 10^{-3}\,\mathrm{K}^{-1}$, and $10^{-2}\,\mathrm{K}^{-1}$.

	CPU Time (s)			Nr Residual Evaluations		
w_0 $(10^{-3}\,\mathrm{K}^{-1})$	2	4.89	10	2	4.89	10
AITK	1.51×10^2	1.73×10^2	1.92×10^2	480	538	569
BRDI	1.20×10^2	1.41×10^2	1.55×10^2	376	417	436
BRDI2	1.18×10^2	1.37×10^2	1.48×10^2	368	407	424
NEWT	NC	2.15×10^2	2.39×10^2	436	781	833
MPE	1.64×10^2	1.81×10^2	1.94×10^2	535	555	580

8.2.2.6 Effect of predictiors

Figure 8.37 presents the effect of employing a linear or quadratic predictor on the number of residual evaluations needed to fully solve the thermomechanical problem under analysis as a function of the thermal expansion coefficient. All methods improve using the polynomial predictors, with the best effects being achieved using the quadratic predictor. The decrease in residual evaluations is around 30% to 40% for all methods using the quadratic predictor, except for the MPE. The polynomial vector extrapolation method displays a much smaller increase in efficiency from the polynomial predictors considered. Finally, the use of predictors allows the Newton-GMRES method to converge for $w_0 = w_h = 2 \times 10^{-3}\,\mathrm{K}^{-1}$.

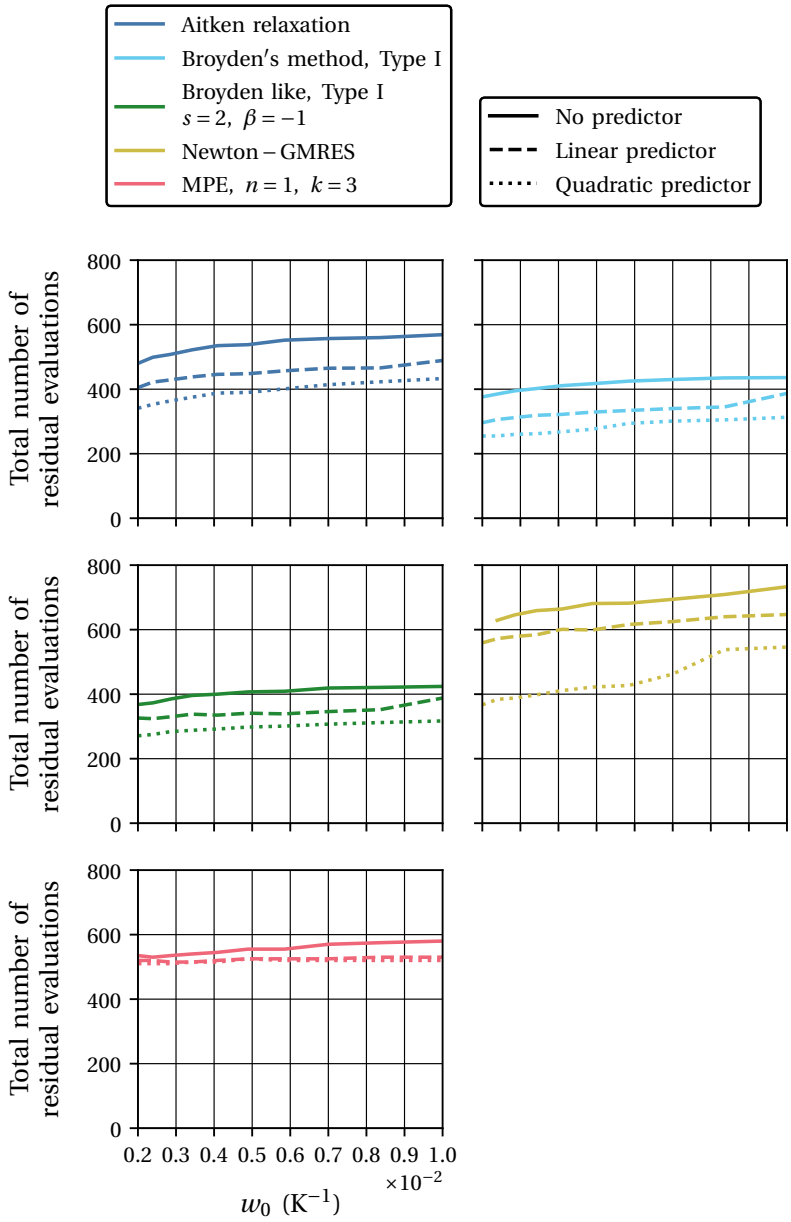


Figure 8.37: Total number of iterations as a function of the thermal softening parameters $w_0 = w_h$ for the best performing implicit methods in each class considered using a linear, a quadratic, and no predictor in the solution of the necking of a circular bar with $w_0 = w_h = 2 \times 10^{-3}\,\mathrm{K}^{-1}$ to $10^{-2}\,\mathrm{K}^{-1}$.

8.3 Conclusions

Given the set of numerical results shown, Broyden-like techniques with $\beta = -1$, $s = 1$, which is also Broyden's method, and $s = 2$, both with Type I update appear to be the most advantageous methods in terms of computational effort. The Aitken relaxation is also effective, particularly for the thermoelastic problem. It requires the least amount of memory and is the easiest to implement. The MPE in cycling mode with $(s,k) = (1,3)$ competes with the Aitken relaxation for the thermo-elastoplastic problem, but is more comparable to the Newton-GMRES with $\eta = 10^{-3}$ for the thermoeleastic problem. Among the techniques deemed the best in each class, the last one performs the poorest. However, because the Jacobian of the residual is well estimated, it is the approach that most easily accepts global strategies such as line search. The requirement for a global strategy isn't illustrated in the problems described, but it may be significant in other circumstances. Last but not least, the polynomial predictors significantly increased computing effectiveness without adding to the complexity of the numerical methods.

Bibliography

Adam, L. and J.-P. Ponthot
2002a. Numerical simulation of viscoplastic and frictional heating during finite deformation of metal. Part I: Theory. *Journal of engineering mechanics*, 128(11):1215–1221. Publisher: American Society of Civil Engineers.

Adam, L. and J.-P. Ponthot
2002b. Numerical simulation of viscoplastic and frictional heating during finite deformation of metal. Part II: Applications. *Journal of engineering mechanics*, 128(11):1222–1232. Publisher: American Society of Civil Engineers.

Adam, L. and J.-P. Ponthot
2005. Thermomechanical modeling of metals at finite strains: First and mixed order finite elements. *International Journal of Solids and Structures*, 42(21):5615–5655.

Agelet de Saracibar, C.
1998. Numerical analysis of coupled thermomechanical frictional contact problems. Computational model and applications. *Archives of Computational Methods in Engineering*, 5(3):243–301.

Agelet de Saracibar, C., M. Cervera, and M. Chiumenti
1999. On the formulation of coupled thermoplastic problems with phase-change. *International Journal of Plasticity*, 15(1):1–34.

Argyris, J. H. and J. S. Doltsinis
1981. On the natural formulation and analysis of large deformation coupled thermomechanical problems. *Computer Methods in Applied Mechanics and Engineering*, 25(2):195–253.

Armero, F.
1999. Formulation and finite element implementation of a multiplicative model of coupled poro-plasticity at finite strains under fully saturated conditions. *Computer Methods in Applied Mechanics and Engineering*, 171:205–241.

Armero, F. and J. C. Simo
1992. A new unconditionally stable fractional step method for non-linear coupled thermomechanical problems. *International Journal for Numerical Methods in Engineering*, 35(4):737–766. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/nme.1620350408.

Armero, F. and J. C. Simo
1993. A priori stability estimates and unconditionally stable product formula

algorithms for nonlinear coupled thermoplasticity. *International Journal of Plasticity*, 9(6):749–782.

Badia, S., A. F. Martín, and R. Planas
2014. Block recursive LU preconditioners for the thermally coupled incompressible inductionless MHD problem. *Journal of Computational Physics*, 274:562–591.

Belytschko, T. and R. Mullen
1976. Mesh partitions of explicit-implicit time integration. *Formulations and computational algorithms in finite element analysis*, Pp. 673–690. Publisher: MIT Press: New York.

Belytschko, T. and R. Mullen
1978. Stability of explicit-implicit mesh partitions in time integration. *International Journal for Numerical Methods in Engineering*, 12(10):1575–1586. Publisher: Wiley Online Library.

Belytschko, T., H.-J. Yen, and R. Mullen
1979. Mixed methods for time integration. *Computer Methods in Applied Mechanics and Engineering*, 17:259–275. Publisher: Elsevier.

Blom, D.
2017. *Efficient numerical methods for partitioned fluid-structure interaction simulations*. Ph.D., Delft University of Technology, Netherlands.

Blom, F. J.
1998. A monolithical fluid-structure interaction algorithm applied to the piston problem. *Computer methods in applied mechanics and engineering*, 167(3-4):369–391. Publisher: Elsevier.

Borja, R., C. Tamagnini, and E. Alarćon
1998. Elastoplastic consolidation at finite strain part 2: finite element implementation and numerical examples. *Computer Methods in Applied Mechanics and Engineering*, 159:103–122.

Brezinski, C. and M. R. Zaglia
2013. *Extrapolation methods: theory and practice*. Elsevier.

Broyden, C. G.
1965. A class of methods for solving nonlinear simultaneous equations. *Mathematics of computation*, 19(92):577–593. Publisher: JSTOR.

Carter, J. P. and J. R. Booker
1989. Finite element analysis of coupled thermoelasticity. *Computers & Structures*, 31(1):73–80.

Causin, P., J.-F. Gerbeau, and F. Nobile
2005. Added-mass effect in the design of partitioned algorithms for fluid–structure problems. *Computer methods in applied mechanics and engineering*, 194(42-44):4506–4527. Publisher: Elsevier.

Çengel, Y. and M. Boles
2011. *Thermodynamics: An Engineering Approach*, Cengel series in engineering thermal-fluid sciences. McGraw-Hill.

Cervera, M., R. Codina, and M. Galindo
1996. On the computational efficiency and implementation of block-iterative algorithms for nonlinear coupled problems. *Engineering Computations*, 13(6):4–30. Publisher: MCB UP Ltd.

Chen, K.
2005. *Matrix Preconditioning Techniques and Applications*, Cambridge Monographs on Applie. Cambridge University Press.

Combescure, A. and A. Gravouil
2002. A numerical scheme to couple subdomains with different time-steps for predominantly linear transient analysis. *Computer Methods in Applied Mechanics and Engineering*, 191(11):1129–1157.

Danowski, C.
2014. *Computational Modelling of Thermo-Structure Interaction with Application to Rocket Nozzles*. Ph.D., Technische Universität München, Germany.

Danowski, C., V. Gravemeier, L. Yoshihara, and W. A. Wall
2013. A monolithic computational approach to thermo-structure interaction. *International Journal for Numerical Methods in Engineering*, 95(13):1053–1078. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/nme.4530.

de Souza Neto, E., D. Peric, and D. Owen
2008. *Computational Methods for Plasticity: Theory and Applications*. Wiley.

Degroote, J.
2010. *Development of algorithms for the partitioned simulation of strongly coupled fluid-structure interaction problems*. PhD Thesis, Ghent University. ISBN: 9789085783442.

Degroote, J., P. Bruggeman, R. Haelterman, and J. Vierendeels
2008. Stability of a coupling technique for partitioned solvers in FSI applications. *Computers & Structures*, 86(23):2224–2234.

Dennis, J. and R. Schnabel
1996. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Classics in Applied Mathematics. Society for Industrial and Applied Mathematics.

Dettmer, W. and D. Perić
2006. A computational framework for fluid–structure interaction: Finite element formulation and applications. *Computer Methods in Applied Mechanics and Engineering*, 195(41):5754–5779.

, DIN EN 1991-1-2/NA:2010-12: Anhang CC 1991. Prüfung und Validierung von Rechenprogramm für Brandschutznachweise mittels allgemeiner Rechenverfahren. Standard, German Institute for Standardization, Berlin, Germany.

Dittmann, M.
2017. *Isogeometric analysis and hierarchical refinement for multi-field contact problems*. Ph.D.

Dittmann, M., M. Franke, İ. Temizer, and C. Hesch
2014. Isogeometric Analysis and thermomechanical Mortar contact problems. *Computer Methods in Applied Mechanics and Engineering*, 274:192–212.

Erbts, P. and A. Düster
2012. Accelerated staggered coupling schemes for problems of thermoelasticity at finite strains. *Computers & Mathematics with Applications*, 64(8):2408–2430.

Erbts, P., S. Hartmann, and A. Düster
2015. A partitioned solution approach for electro-thermo-mechanical problems. *Archive of Applied Mechanics*, 85(8):1075–1101.

Fang, H.-r. and Y. Saad
2009. Two classes of multisecant methods for nonlinear acceleration. *Numerical linear algebra with applications*, 16(3):197–221. Publisher: Wiley Online Library.

Farhat, C., P. Geuzaine, and G. Brown
2003. Application of a three-field nonlinear fluid–structure formulation to the prediction of the aeroelastic parameters of an F-16 fighter. *Computers & Fluids*, 32:3–29.

Farhat, C. and M. Lesoinne
2000. Two efficient staggered algorithms for the serial and parallel solution of three-dimensional nonlinear transient aeroelastic problems. *Computer methods in applied mechanics and engineering*, 182(3-4):499–515. Publisher: Elsevier.

Farhat, C., M. Lesoinne, P. Stern, and S. Lanteri
1997. High performance solution of three-dimensional nonlinear aeroelastic problems via parallel partitioned algorithms: methodology and preliminary results. *Advances in Engineering Software*, 28:43–61.

Farhat, C., K. Park, and Y. Dubois-Pelerin
1991. An unconditionally stable staggered algorithm for transient finite element analysis of coupled thermoelastic problems. *Applied Mechanics and Engineering*, 85:349–365.

Farhat, C., A. Rallu, K. G. Wang, and T. Belytschko
2010. Robust and provably second-order explicit-explicit and implicit-explicit staggered time-integrators for highly non-linear compressible fluid-structure interaction problems. *International Journal for Numerical Methods in Engineering*, 84:73–107.

Farhat, C., K. Zee, and P. Geuzaine
2006. Provably second-order time-accurate loosely-coupled solution algorithms for transient nonlinear computational aeroelasticity. *Computer Methods in Applied Mechanics and Engineering*, 195:1973–2001.

Felder, S., N. Kopic-Osmanovic, H. Holthusen, T. Brepols, and S. Reese
2021. Thermo-mechanically coupled gradient-extended damage-plasticity modeling of metallic materials at finite strains. *International Journal of Plasticity*, P. 103142.

Felippa, C. and T. L. Geers
1988. Partitioned analysis for coupled mechanical systems. *Engineering Computations*, 5:123–133.

Felippa, C. A. and K. C. Park
1980. Staggered transient analysis procedures for coupled mechanical systems: Formulation. *Computer Methods in Applied Mechanics and Engineering*, 24(1):61–111.

Felippa, C. A., K. C. Park, and C. Farhat
2001. Partitioned analysis of coupled mechanical systems. *Computer Methods in Applied Mechanics and Engineering*, 190(24):3247–3270.

Fernández, M. A., J.-F. Gerbeau, and C. Grandmont
2006. A projection semi-implicit scheme for the coupling of an elasticstructure with an incompressible fluid. *International Journal for Numerical Methods in Engineering*.

Förster, C.
2007. Robust methods for fluid-structure interaction with stabilised finite elements.

Förster, C., W. A. Wall, and E. Ramm
2007. Artificial added mass instabilities in sequential staggered coupling of nonlinear structures and incompressible viscous flows. *Computer methods in applied mechanics and engineering*, 196(7):1278–1293. Publisher: Elsevier.

Gatzhammer, B.
2014. *Efficient and Flexible Partitioned Simulation of Fluid-Structure Interactions*. Dissertation, Technische Universität München, München.

Gee, M. W., U. Küttler, and W. A. Wall
2011. Truly monolithic algebraic multigrid for fluid–structure interaction. *International Journal for Numerical Methods in Engineering*, 85(8):987–1016. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/nme.3001.

Ghadiani, S. R.
2005. *A multiphasic continuum mechanical model for design investigations of an effusion-cooled rocket thrust chamber*.

Gillard, J.
2019. *An Efficient Partitioned Coupling Scheme for Tire Hydroplaning Analysis*. Ph.D., Technishcen Universität München, München.

Gitterle, M.
2012. *A dual mortar formulation for finite deformation frictional contact problems including wear and thermal coupling*. Ph.D., Technische Universität München.

Glaser, S.
1992. *Gekoppelte thermomechanische Berechnung duennwandiger Strukturen mit der Methode der Finiten Elemente*. Ph.D., Institute fuer Statik und Dynamik der Luft- und Raumfahrtkonstruktionen, University of Stuttgart.

Haelterman, R., J. Degroote, D. Van Heule, and J. Vierendeels
2009. The Quasi-Newton Least Squares Method: A New and Fast Secant Method Analyzed for Linear Systems. *SIAM Journal on Numerical Analysis*, 47(3):2347–2368. _eprint: https://doi.org/10.1137/070710469.

Hansen, G.
2011. A Jacobian-free Newton Krylov method for mortar-discretized thermomechanical contact problems. *Journal of Computational Physics*, 230(17):6546–6562.

Heil, M.
2004. An efficient solver for the fully coupled solution of large-displacement fluid–structure interaction problems. *Computer Methods in Applied Mechanics and Engineering*, 193(1):1–23.

Hesch, C. and P. Betsch
2011. Energy-momentum consistent algorithms for dynamic thermomechanical problems—Application to mortar domain decomposition problems. *International Journal for Numerical Methods in Engineering*, 86(11):1277–1302. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/nme.3095.

Holt, M. and N. Yanenko
2012. *The Method of Fractional Steps: The Solution of Problems of Mathematical Physics in Several Variables*. Springer Berlin Heidelberg.

Holzapfel, G. A.
2000. *Nonlinear Solid Mechanics: A Continuum Approach for Engineering*. Wiley. Google-Books-ID: _ZkeAQAAIAAJ.

Holzapfel, G. A. and J. C. Simo
1996. Entropy elasticity of isotropic rubber-like solids at finite strains. *Computer Methods in Applied Mechanics and Engineering*, 132(1):17–44.

Hron, J. and S. Turek
2006. A Monolithic FEM/Multigrid Solver for an ALE Formulation of Fluid-Structure Interaction with Applications in Biomechanics. In *Fluid-Structure Interaction.*, Lecture Notes in Computational Science and Engineering, vol 53.

Hübner, B., E. Walhorn, and D. Dinkler
2004. A monolithic approach to fluid–structure interaction using space–time finite elements. *Computer Methods in Applied Mechanics and Engineering*, 193(23):2087–2104.

Hüeber, S. and B. I. Wohlmuth
2009. Thermo-mechanical contact problems on non-matching meshes. *Computer Methods in Applied Mechanics and Engineering*, 198(15):1338–1350.

Hughes, T. J. and W. Liu
1978. Implicit-explicit finite elements in transient analysis: stability theory.

Ibrahimbegovic, A.
2009. Thermodynamics and solution methods for coupled problems. In *Nonlinear Solid Mechanics*, volume 160, Pp. 427–474. Dordrecht: Springer Netherlands. Series Title: Solid Mechanics and its Applications.

Ibrahimbegovic, A. and L. Chorfi
2002. Covariant principal axis formulation of associated coupled thermoplasticity at finite strains and its numerical implementation. *International Journal of Solids and Structures*, 39(2):499–528.

Irons, B. M. and R. C. Tuck
1969. A version of the Aitken accelerator for computer iteration. *International Journal for Numerical Methods in Engineering*, 1(3):275–277. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/nme.1620010306.

Jha, B. and R. Juanes
2007. A locally conservative finite element framework for the simulation of coupled flow and reservoir geomechanics. *Acta Geotechnica*, 2:139–153.

Johansson, L. and A. Klarbring
1993. Thermoelastic frictional contact problems: Modelling, finite element approximation and numerical realization. *Computer Methods in Applied Mechanics and Engineering*, 105(2):181–210.

Joosten, M. M., W. G. Dettmer, and D. Perić
2009. Analysis of the block Gauss–Seidel solution procedure for a strongly coupled model problem with reference to fluid–structure interaction. *International Journal for Numerical Methods in Engineering*, 78(7):757–778. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/nme.2503.

Kelley, C.
2003. *Solving Nonlinear Equations with Newton's Method*, Fundamentals of Algorithms. Society for Industrial and Applied Mathematics.

Kim, J., H. Tchelepi, and R. Juanes
2011a. Stability and convergence of sequential methods for coupled flow and geomechanics: Drained and undrained splits. *Computer Methods in Applied Mechanics and Engineering*, 200:2094–2116.

Kim, J., H. Tchelepi, and R. Juanes
2011b. Stability and convergence of sequential methods for coupled flow and geomechanics: Fixed-stress and fixed-strain splits. *Computer Methods in Applied Mechanics and Engineering*, 200:1591–1606.

Klöppel, T., A. Popp, U. Küttler, and W. A. Wall
2011. Fluid–structure interaction for non-conforming interfaces based on a dual mortar formulation. *Computer Methods in Applied Mechanics and Engineering*, 200(45):3111–3126.

Küttler, U., M. Gee, C. Förster, A. Comerford, and W. A. Wall
2010. Coupling strategies for biomedical fluid–structure interaction problems. *International Journal for Numerical Methods in Biomedical Engineering*, 26(3-4):305–321. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cnm.1281.

Küttler, U. and W. A. Wall
2008. Fixed-point fluid–structure interaction solvers with dynamic relaxation. *Computational Mechanics*, 43(1):61–72.

Küttler, U. and W. A. Wall
2009. Vector Extrapolation for Strong Coupling Fluid-Structure Interaction Solvers. *Journal of Applied Mechanics*, 76(2).

Lenarda, P. and M. Paggi
2016. A geometrical multi-scale numerical method for coupled hygro-thermo-mechanical problems in photovoltaic laminates. *Computational Mechanics*, 57(6):947–963.

Lewis, R. and Y. Sukirman
1993. Finite element modelling of three-phase flow in deforming saturated oil reservoirs. *International Journal for Numerical and Analytical Methods in Geomechanics*, 17(8):577–598. Publisher: Wiley Online Library.

Lin, P. T., J. N. Shadid, R. S. Tuminaro, M. Sala, G. L. Hennigan, and R. P. Pawlowski
2010. A parallel fully coupled algebraic multilevel preconditioner applied to multiphysics PDE applications: Drift-diffusion, flow/transport/reaction, resistive MHD. *International Journal for Numerical Methods in Fluids*, 64(10-12):1148–1179. tex.eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/fld.2402.

Matthies, H. and J. Steindorf
2003a. Partitioned Strong Coupling Algorithms for Fluid-Structure-Interaction. *Computers & Structures*, 81:805–812.

Matthies, H. and J. Steindorf
2003b. Strong Coupling Methods.

Mayr, M., T. Klöppel, W. A. Wall, and M. W. Gee
2015. A Temporal Consistent Monolithic Approach to Fluid-Structure Interaction Enabling Single Field Predictors. *SIAM Journal on Scientific Computing*, 37(1):B30–B59. Publisher: Society for Industrial and Applied Mathematics.

Mayr, M., M. H. Noll, and M. W. Gee
2020. A hybrid interface preconditioner for monolithic fluid–structure interaction solvers. *Advanced Modeling and Simulation in Engineering Sciences*, 7(1):15.

Michler, C.
2005. *Efficient numerical methods for fluid-structure interaction*. Ph.D., Delft University of Technology, Netherlands. ISBN: 90-9019533-5.

Michler, C., E. H. v. Brummelen, S. J. Hulshoff, and R. d. Borst
2003. The relevance of conservation for stability and accuracy of numerical methods for fluid–structure interaction. *Computer Methods in Applied Mechanics and Engineering*, 192(37):4195–4215.

Michler, C., S. Hulshoff, E. Van Brummelen, and R. De Borst
2004. A monolithic approach to fluid–structure interaction. *Computers & fluids*, 33(5-6):839–848. Publisher: Elsevier.

Michler, C., E. H. van Brummelen, and R. de Borst
2005. An interface Newton–Krylov solver for fluid–structure interaction. *International Journal for Numerical Methods in Fluids*, 47(10-11):1189–1195. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/fld.850.

Miehe, C.
1995a. Entropic thermoelasticity at finite strains. Aspects of the formulation and numerical implementation. *Computer Methods in Applied Mechanics and Engineering*, 120(3):243–269.

Miehe, C.
1995b. A theory of large-strain isotropic thermoplasticity based on metric transformation tensors. *Archive of Applied Mechanics*, 66(1):45–64.

Mikelić, A. and M. F. Wheeler
2013. Convergence of iterative coupling for coupled flow and geomechanics. *Computational Geosciences*, 17(3):455–461. Publisher: Springer.

Miller, B. A.
2015. *Loosely Coupled Time Integration of Fluid- Thermal-Structural Interactions in Hypersonic Flows*. Ph.D., Ohio State University.

Neishlos, H.
1983. Finite-Element Mesh Partitioning for Time Integration of Transient Problems. In *Numerical Solution of Partial Differential Equations: Theory, Tools and Case Studies: Summer Seminar Series Held at CSIR, Pretoria, February 8–10, 1982*, D. P. Laurie, ed., Pp. 225–245. Basel: Birkhäuser Basel.

Netz, T.
2013. *High-order space and time discretization scheme applied to problems of finite thermo-viscoelasticity*. Ph.D., Institute of Applied Mechanics, Clausthal University of Technology.

Novascone, S. R., B. W. Spencer, J. D. Hales, and R. L. Williamson
2015. Evaluation of coupling approaches for thermomechanical simulations. *Nuclear Engineering and Design*, 295:910–921.

Oancea, V. G. and T. A. Laursen
1997. A finite element formulation of thermomechanical rate-dependent frictional sliding. *International Journal for Numerical Methods in Engineering*, 40(23):4275–4311. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/%28SICI%291097-0207%2819971215%2940%3A23%3C4275%3A%3AAID-NME257%3E3.0.CO%3B2-K.

Pantuso, D., K.-J. Bathe, and P. A. Bouzinov
2000. A finite element procedure for the analysis of thermo-mechanical solids in contact. *Computers & Structures*, 75(6):551–573.

Park, K.
1983. Stabilization of partitioned solution procedure for pore fluid-soil interaction analysis. *International Journal for Numerical Methods in Engineering*, 19:1669–1673.

Park, K., C. Felippa, and J. DeRuntz
1977. Stabilization of staggered solution procedures for fluid-structure interaction analysis. *Computational methods for fluid-structure interaction problems*, 26(94-124):51. Publisher: ASME New York.

Piperno, S.
1997. Explicit/implicit fluid/structure staggered procedures with a structural predictor and fluid subcycling for 2D inviscid aeroelastic simulations. *International Journal for Numerical Methods in Fluids*, 25:1207–1226.

Piperno, S. and C. Farhat
2001. Partitioned procedures for the transient solution of coupled aeroelastic

problems–Part II: energy transfer analysis and three-dimensional applications. *Computer methods in applied mechanics and engineering*, 190(24-25):3147–3170. Publisher: Elsevier.

Piperno, S., C. Farhat, and B. Larrouturou
1995. Partitioned procedures for the transient solution of coupled aroelastic problems Part I: Model problem, theory and two-dimensional application. *Computer Methods in Applied Mechanics and Engineering*, 124:79–112.

Rothe, S., P. Erbts, A. Düster, and S. Hartmann
2015. Monolithic and partitioned coupling schemes for thermo-viscoplasticity. *Computer Methods in Applied Mechanics and Engineering*, 293:375–410.

Saetta, A. and R. Vitaliani
1992. Unconditionally convergent partitioned solution procedure for dynamic coupled mechanical systems. *International Journal for Numerical Methods in Engineering*, 33:1975–1996.

Scheufele, K.
2018. *Coupling schemes and inexact Newton for multi-physics and coupled optimization problems.* PhD Thesis, Universität Stuttgart, Stuttgart.

Seitz, A.
2019. *Computational Methods for Thermo-Elasto-Plastic Contact.* Ph.D., Technische Universität München, Germany.

Seitz, A., W. A. Wall, and A. Popp
2018. A computational approach for thermo-elasto-plastic frictional contact based on a monolithic formulation using non-smooth nonlinear complementarity functions. *Advanced Modeling and Simulation in Engineering Sciences*, 5(1):5.

Shadid, J., R. Pawlowski, J. Banks, L. Chacón, P. Lin, and R. Tuminaro
2010. Towards a scalable fully-implicit fully-coupled resistive MHD formulation with stabilized FE methods. *Journal of Computational Physics*, 229(20):7649–7671.

Sidi, A.
2017. *Vector extrapolation methods with applications.* SIAM.

Simo, J. C. and F. Armero
1992. Recent Advances in the Numerical Analysis and Simulation of Thermoplasticity at Finite Strains.

Simo, J. C. and C. Miehe
1992. Associative coupled thermoplasticity at finite strains: Formulation, numerical analysis and implementation. *Computer Methods in Applied Mechanics and Engineering*, 98(1):41–104.

Smith, B., P. Bjorstad, and W. Gropp
2004. *Domain Decomposition: Parallel Multilevel Methods for Elliptic Partial Differential Equations.* Cambridge University Press.

Tezduyar, T. E., S. Sathe, R. Keedy, and K. Stein
2006. Space–time finite element techniques for computation of fluid–structure interactions. *Computer methods in applied mechanics and engineering*, 195(17-18):2002–2027. Publisher: Elsevier.

, The Standard NAFEMS benchmarks. NAFEMS P18. Publication, NAFEMS.

Torii, R., M. Oshima, T. Kobayashi, K. Takagi, and T. E. Tezduyar
2006. Computer modeling of cardiovascular fluid–structure interactions with the deforming-spatial-domain/stabilized space–time formulation. *Computer Methods in Applied Mechanics and Engineering*, 195(13-16):1885–1895. Publisher: Elsevier.

Traub, J. F.
1982. *Iterative methods for the solution of equations*, volume 312. American Mathematical Soc.

Uekermann, B. W.
2016. *Partitioned fluid-structure interaction on massively parallel systems*. PhD Thesis, Technische Universität München.

Verdugo, F. and W. A. Wall
2016. Unified computational framework for the efficient solution of n-field coupled problems with monolithic schemes. *Computer Methods in Applied Mechanics and Engineering*, 310:335–366.

Vierendeels, J., L. Lanoye, J. Degroote, and P. Verdonck
2007. Implicit coupling of partitioned fluid–structure interaction problems with reduced order models. *Computers & structures*, 85(11-14):970–976. Publisher: Elsevier.

Wall, W. A., S. Genkinger, and E. Ramm
2007. A strong coupling partitioned approach for fluid–structure interaction with free surfaces. *Computers & Fluids*, 36(1):169–183.

Wendt, G., P. Erbts, and A. Düster
2015. Partitioned coupling strategies for multi-physically coupled radiative heat transfer problems. *Journal of Computational Physics*, 300:327–351.

White, J. A. and R. Borja
2008. Stabilized low-order finite elements for coupled solid-deformation/fluid-diffusion and their application to fault zone transients. *Computer Methods in Applied Mechanics and Engineering*, 197:4353–4366.

Wriggers, P., C. Miehe, M. Kleiber, and J. C. Simo
1992. On the coupled thermomechanical treatment of necking problems via finite element methods. *International Journal for Numerical Methods in Engineering*, 33(4):869–883. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/nme.1620330413.

Zavarise, G., P. Wriggers, E. Stein, and B. A. Schrefler
1992. Real contact mechanisms and finite element formulation—a coupled thermomechanical approach. *International Journal for Numerical Methods in Engineering*, 35(4):767–785. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/nme.1620350409.

Zhang, Q. and T. Hisada
2004. Studies of the strong coupling and weak coupling methods in FSI analysis.

International Journal for Numerical Methods in Engineering, 60(12):2013–2029. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/nme.1034.

Zienkiewicz, O., D. K. Paul, and A. Chan
1988. Unconditionally stable staggered solution procedure for soil-pore fluid interaction problems. *International Journal for Numerical Methods in Engineering*, 26:1039–1055.