

Computational Weld Mechanics (CWM) Framework for Exploring Parametric Design Space to Manage Weld Optimization

by

Mahyar Asadi

A Dissertation submitted to
the Faculty of Graduate Studies and Research
in partial fulfilment of
the requirements for the degree of
Doctor of Philosophy

Ottawa-Carleton Institute for
Mechanical and Aerospace Engineering

Department of Mechanical and Aerospace Engineering
Carleton University
Ottawa, Ontario, Canada
April 2011



Library and Archives
Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file *Votre référence*
ISBN: 978-0-494-81543-4
Our file *Notre référence*
ISBN: 978-0-494-81543-4

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.


Canada

Copyright ©
2011 - Mahyar Asadi

The undersigned recommend to
the Faculty of Graduate Studies and Research
acceptance of the Dissertation

**Computational Weld Mechanics (CWM) Framework for
Exploring Parametric Design Space to Manage Weld
Optimization**

Submitted by **Mahyar Asadi**
in partial fulfilment of the requirements for the degree of
Doctor of Philosophy

John A. Goldak, Supervisor

Mettin Yaras, Department Chair

Carleton University

2011

Abstract

The goal is to lead the development of a computational weld mechanics (CWM) framework that automates multiple setups and evaluations required to explore a design space by given design of experiment (DOE) matrices. Saving an expert-user's time to prepare several analyses and allocating CPUs to be utilized efficiently make this framework cost effective and time effective to manage designer-driven optimization and control application of CWM. A validation analysis is conducted in this framework to minimize the control vectors of CWM. Actual CWM problems with continuous and/or discontinuous parametric design spaces are solved in this framework to minimize weld distortion using derivative-free optimization algorithms and DOE that become attractive in this framework.

I would like to dedicate this thesis to my wife, Sharareh
and my parents, Pouran and Hashem.

Acknowledgments

I owe my knowledge and deepest gratitude to my supervisor Dr. John Goldak who has been my mentor and advisor with endless source of knowledge and experience. I am very thankful for his patient guidance, encouragement and excellent support since the first time I saw him. I am very grateful for the financial support I have received.

It is a pleasure to thank those who made this thesis possible; Daniel Downey, Stanislav Tchernov and Jianguo Zhou. I would like to show my gratitude to Dr. Hughes, Dr. Nielsen and Dr. Fong for their kind advice.

Most of all, I take this opportunity to express my profound gratitude to my wife, Sharareh, for her tremendous encouragement and support in all aspects of my life. Her role is definitely indispensable and cherished in my achievements. I am indebted forever to my parents, Pouran and Hashem for their unconditional support and I would like to thank my brothers Mazyar and Kamyar, my parents-in-law Zahra and Hossein and sister-in-law Shahrzad for their moral supports.

Preface

Motivation of this study is presented in the first chapter by looking to the limitations of practical applications of CWM. Chapter 2 discusses the supports required to implement optimization in the design stage from a designer point of view and a framework is suggested. The framework uses DOE matrices as a communication language with a designer. The optimization and control methods that helps to create the DOE matrices, are reviewed in Appendix B as a complement of chapter 2. Chapter 4 presents the experience gained in conducting a CWM validation test and related challenges. To develop such a framework, actual CWM problems were solved including design space discretization DOE matrix (chapter 5) for parametric design space exploration, regular and least-square direct search (chapter 6), regression analysis and Monte Carlo (chapter 7), control and sensitivity (chapter 8), and a surrogate model (chapter 9) to find an optimal design in the design space of different mitigation techniques, residual stress and combinatorial weld sequence. There is a comparison among them in chapter 10 in terms of DOE matrices and CPU time. Each CWM problem solved presents its own conclusions in the context of welding process at the end of each chapter. Overall conclusions and suggested future work are presented in the last chapter.

Table of Contents

Abstract	iv
Acknowledgments	vi
Preface	vii
Table of Contents	viii
List of Tables	xiv
List of Figures	xvii
Nomenclature	xxix
1 Computational Weld Mechanics	1
1.1 CWM Control and Optimization	3
1.2 Industrial CWM Problems	6
2 Control and Optimization Problems	8
2.1 Objective Function	9
2.2 Optimization Methods	9
2.3 Parametric Design Space	10
2.4 CWM Optimizer Framework	14

3	CWM Design Space Exploration Using DOE Matrices	18
3.1	Validation of CWM	18
3.2	Design Space Discretization	19
3.3	Regular and Least-Square Direct-Search	20
3.4	Regression/Response Surface Analysis	21
3.5	Control and Sensitivity Analysis	21
3.6	Surrogate/Meta Model	22
3.7	Comparison of DOE Matrices Used in the Projects	23
3.8	Additional Projects	23
3.8.1	Optimizing Temper Bead Welding by Computational Weld Me- chanics and Design of Experiment Matrix	24
3.8.2	Power per Unit Length of Weld Does not Characterize a Weld	24
3.8.3	Microstructure and Hardness Modelling: Effect of Multiple Bead Deposition in Temper Bead Welding Technique	25
4	Challenges in Validation of CWM Software to Compute Residual Stress and Distortion in Welds	26
4.1	Analysis Background	26
4.2	Computational Weld Mechanics as a Control Problem	28
4.2.1	Guidelines for Verification and Validation for Use in an Appli- cation	30
4.2.2	Verification of VrWeld	31
4.2.3	Validation of VrWeld	33
4.3	Validation Process	34
4.3.1	Validation of Transient Thermal Analysis	35
4.3.2	Validation of Microstructure Evolution	36
4.3.3	Validation of Transient Strain and Displacement Analyses . .	37

4.3.4	Sensitivity of Computed Results to Model Parameters.	37
4.4	A Validation Example	38
4.4.1	Validation of Transient Temperature Computation	39
4.4.2	Validation of Transient Strain Gauge Computation	40
4.5	Closure of the Analysis	42
5	A Continuous Design Space Exploration; Space Discretization	46
5.1	Analysis Background	47
5.2	Computational Setups	48
5.2.1	Thermal Analysis	49
5.2.2	Stress Analysis	49
5.3	Optimization Problem	49
5.3.1	Prescribed Displacement in the Middle	51
5.3.2	Parabolic Displacement at the Bottom	55
5.4	CWM Results and Comparison	59
5.5	Closure of the Analysis	60
6	A Continuous Design Space Exploration; Direct Searching in the Space	63
6.1	Analysis Background	64
6.2	Direct-Search Algorithm	65
6.3	Least-Square Algorithm	65
6.4	Edge-Welded Bar Setup	67
6.5	Computational Setups	71
6.5.1	Thermal Analysis	71
6.5.2	Stress Analysis	71
6.6	Optimization Problem for the Edge-Welded Bar	72
6.6.1	Original Direct-Search Algorithm	75

6.6.2	Least-Square Direct-Search Algorithm	75
6.7	Comparison and Discussion	77
6.8	CWM Results for the Optimum	81
6.9	Closure of the Analysis	82
7	A Continuous Design Space Exploration; Regression Modelling	84
7.1	Analysis Background	84
7.2	Experimental Data	86
7.3	Computational Setups	87
7.4	Comparison	91
7.5	Uncertainty Analysis	94
7.6	Perturbation Analysis	100
7.7	Closure of the Analysis	103
8	A Continuous Design Space Exploration; Control and Sensitivity	
	Analysis	105
8.1	Analysis Background	106
8.2	Computational Setup	107
8.2.1	Thermal Analysis Setup	109
8.2.2	Stress Analysis Setup	110
8.3	Transverse Traction Sampling	110
8.4	Power per Unit Length Sampling	114
8.5	Double Ellipsoid Adjustment	114
8.6	Hot Crack Sensitivity Analyses	127
8.7	Discussion	128
8.8	Closure of the Analysis	141

9 A Discontinuous Combinatorial Design Space Exploration; Surrogate Modelling	144
9.1 Analysis Background	145
9.2 Girth Weld in Pipe	147
9.3 Surrogate Model	149
9.4 Computational Setups	153
9.4.1 Thermal simulation	153
9.4.2 Stress simulation	153
9.5 Surrogate Algorithm for Girth Weld	156
9.6 Surrogate Results and Comparison with Exact Results	165
9.7 Find the Optimum	173
9.8 Compare the Best and Worst	173
9.9 Closure of the Analysis	175
10 Designer-Driven Exploration of CWM Design Space	182
10.1 Problem Background	182
10.2 Optimization Problems and Comparison	183
11 Conclusion	186
11.1 Looking to the Future	187
Appendix A Multi-Physics of Welding	196
A.1 Thermal Analysis in Welding	197
A.1.1 Heat Source Models	198
A.2 Microstructure of Low-Alloy Steels	201
A.2.1 Hardness Evaluation for Low-Alloy Steels	202
A.3 Stress Analysis in Welding	203
List of References	196

Appendix B Optimization Methods	205
B.1 Derivative-Based Optimization	205
B.1.1 Finite Difference	206
B.1.2 Direct Differentiation	206
B.1.3 Adjoint Methods	208
B.1.4 Derivative-Based Optimization Drawbacks	211
B.2 Derivative-Free Optimization	211
B.2.1 Nelder-Mead Downhill Simplex	212
B.2.2 Direct-Search Algorithms	213
B.2.3 Least-Square Algorithm	214
B.2.4 Line Search Algorithms	216
B.2.5 Surrogate/Meta Model Algorithms	220

List of Tables

5.1	DOE matrix employed for prescribed displacement in the middle of the bar.	52
5.2	DOE matrix employed for parabolic prescribed displacement.	57
6.1	L_9 Taguchi DOE matrix employed to screen the optimization parameters wrt the sensitivity of the objective function to each parameters.	73
6.2	DOE matrix used to minimize the objective function, wrt to the radius and power respecting the constraint.	74
6.3	Data table for the original direct-search algorithm tested and the decision at the end of each iteration. C indicates the result copied from previous runs to avoid repeating the analysis.	76
6.4	Data table for the least-square direct-search algorithm tested and the decision at the end of each iteration. C indicates the result copied from previous runs to avoid repeating the analysis.	78
6.5	Total CPU time and the number of cores used for each iteration in the original direct-search DOE matrix.	80
6.6	Total CPU time and the number of cores used for each iteration in the least-square direct-search DOE matrix.	80
7.1	Chemical composition of parent and consumable materials [wt%]	86
7.2	Design matrix, $X_{81 \times 15}$	97
7.3	Values of each level of double ellipsoid parameters [mm]	98

7.4	Response matrix scheme $Y_{81 \times 78}$	98
7.5	Coefficient matrix scheme $\hat{\beta}_{15 \times 78}$	99
8.1	Process Prameters	109
8.2	Transverse traction and corresponding stresses applied to the side surfaces.	114
8.3	Power and speed tested for constant power per unit length.	116
8.4	Adjusted double ellipsoid parameters.	116
8.5	Variation for F, R, and D.	117
8.6	DOE matrix of all combinations for variation given in Table 8.5 to the references F, R, and D.	118
8.7	Second iteration of variation in F, R, and D for project 1.	122
8.8	Second iteration of variation in F, R, and D for project 7.	123
8.9	DOE matrix generated from the second variation given in Table 8.7 for project 1.	124
8.10	DOE matrix generated from the second variation given in Table 8.7 for project 7. '0' denotes the minimum double ellipsoid size of 0.0002 [m].	125
9.1	Weld procedure used for each layer of weld in this analysis.	148
9.2	All possible sequences for the first layer.	159
9.3	Maximum displacement for the first layer [mm].	159
9.4	All sequences for the second layer.	162
9.5	Sequences analyzed for the second layer.	163
9.6	The R_2'' -matrix.	163
9.7	The R_1' -matrix.	164
9.8	List of choices for sequence (c, a, b).	168
9.9	List of choices for sequence (b, c, a).	169
9.10	List of choices for sequence (-a, b, c).	171

10.1 DOE matrix sizes, total CPU times, and number of cores used in the project discussed in chapters 5, 6, 7, 8 and 9.	184
---	-----

List of Figures

2.1	Suggested optimizer framework	17
4.1	This figure taken from [22] provides an over-view of the verification and validation process. The left side is devoted to the development of the model. The right side is devoted to the development of experimental data.	32
4.2	Specimen and locations of sensors and gauges [27].	38
4.3	Comparison between computed thermal profiles and measured profiles for the four thermocouples. The red and green curves are experimental and computed temperatures for thermocouple 1, the blue and magenta curves are experimental and computed temperatures for thermocouple 2, the blue-green and yellow curves are experimental and computed temperatures for thermocouple 3, the black and orange curves are experimental and computed temperatures for thermocouple 4.	39
4.4	Comparison between the computed result for strain gauges with the experimental data taken from [27]. Thermal strain is fully subtracted from the total strain. Red, dark blue, light blue and black are experimental data for strain gauge 4, 3, 2, and 1 respectively. Green, magenta, yellow and light brown are computed results for strain gauge 4, 3, 2, and 1 respectively.	40

4.5	Sensitivity of strain gauge 4 to its position distanced from top edge by 4.13, 3.81, 3.49, 3.17, 2.85 [mm] are shown in green, dark blue, magenta, light blue and yellow respectively. Experimental data is shown in red.	41
4.6	Comparison of experimental measured transient displacement, shown in red, with the computed transient displacement. Green shows the setting of cut-off temperature to 700 [$^{\circ}K$], convergence criteria 0.001 and maximum number of NR iterations to 1. Blue shows the setting of cut off temperature to 700 [$^{\circ}K$], convergence criteria 0.000001, maximum number of NR iterations to 10. Magenta shows the setting of cut off temperature to 800 [$^{\circ}K$], convergence criteria 0.000001, maximum number of NR iterations to 10. Light blue shows the setting of cut off temperature to 840 [$^{\circ}K$], convergence criteria 0.000001, maximum number of NR iterations to 10. Yellow shows the setting of cut off temperature to 840 [$^{\circ}K$], convergence criteria 0.000001, maximum number of NR iterations to 10 and an increase in CTE of about 10 %.	43
4.7	Comparison the measured mid-length deflection and computed mid-length deflection with setting of cut-off temperature to 850 [$^{\circ}K$] (melting temperature 880 [$^{\circ}K$]), convergence criteria to 1e-6, maximum number of NR iterations to 10 and an increased in CTE by adding 2e-6 to the original CTE values taken from [27].	44
5.1	Final distortion (50x) when the part is free to deform but rigid body modes are constrained to zero. Red and Yellow lines show the positive X and Y directions.	50

5.2	Objective function, i.e. maximum final Y -displacement at the end of the process. Each curve illustrates a fitted curve for one prescribed displacement for the range of delay time. Data is given in Table 5.1 and the fitted curve equations are given in Eq. 5.1, 5.2, 5.3, and 5.4 respectively.	53
5.3	Final Y -displacement plane vs the prescribed displacement and the delay time achieved for the nodal prescribed displacement at the end of process and the zero final displacement plane intersects it.	54
5.4	The curve shows the prescribed displacement and the delay time required to achieve zero final displacement. The equation is given in 5.5.	54
5.5	Objective function, i.e. maximum final Y -displacement at the end of the process. Each curve illustrates a fitted curve for each h values. Data is given in Table 5.2 and the fitted curve equations are given in Eq. 5.7, Eq. 5.8, Eq. 5.9, and Eq. 5.10 respectively.	58
5.6	Final Y -displacement plane vs h and the delay time achieved for the parabolic prescribed displacement at the end of process and the zero final displacement plane intersects it.	58
5.7	Curve shows h (mm) and the delay time required to achieve zero final displacement. The equation is given in 5.11.	59
5.8	Final deflection from the two mitigation strategies employed. Distance is from the left bottom corner to the right bottom corner of the bar. Units are [m].	61
5.9	Longitudinal residual stress, σ_{xx} , plotted along a line normal to the weld, Y -direction, from the top edge to the bottom edge of the bar in the mid-length. Units are [Pa] and [m].	61
6.1	A 2D view of the 3D mesh employed in the analysis.	68

6.2	Deflection in Y direction at the end of the process (x85). Red is X direction and yellow is Y direction.	68
6.3	Origin of the coordinate system used in the analysis. Red is X direction and yellow is Y direction.	70
6.4	Constraint showing the allowable region for two side heater parameters: power and area. Nodes on the gray zone have the maximum temperature in a side heater below 480 [°K] and therefore generate no plastic strain.	70
6.5	The original direct-search algorithm results (Table 6.3) is illustrated graphically to show the path followed by this method to the minimum.	77
6.6	The least-square direct-search algorithm results (Table 6.4) are illustrated in blue. It is compared to the original direct-search in red (Table 6.3) to show the path followed by either algorithms to the minimum.	79
6.7	Final deflection [m] for the weld with no mitigation, with the side heater only applied, and with 2the weld mitigated by the side heater. Distance is from the left bottom corner to the right bottom corner of the bar. Units are [m].	81
6.8	Longitudinal residual stress, σ_{xx} , in the bar after welding is complete for the weld with no mitigation, when the side heater only applied and the weld mitigated by the side heater.	82
7.1	Experimental setup that shows the filler metal on the 200x100x12 [mm] plate and the line along which the elastic strain profile is measured .	87
7.2	Thermal results showing transient temperatures with filler metal being added, after 3.5 seconds, top, and 25 seconds, bottom.	88
7.3	Microstructure analysis including alpha or ferrite phase, top, and pearlite phase, bottom.	89

7.4	Microstructure analysis including martensite phase, top, and hardness distribution, bottom.	90
7.5	Total displacement (left) and effective stress (right) in the part after welding, cool down to room temperature and then releasing the restraints.	92
7.6	Residual stress profile for longitudinal, transverse and normal stress components (see Fig. 7.1) on the same path of the experimental measurement are shown in Fig. 7.1. The solid and dotted lines designate computed and experimental results. Colors; red, black and light blue show the computed longitudinal, transverse and normal components and dark blue, pink and green show the experimental ones respectively.	93
7.7	The difference between the computed and measured values at each node. Blue is longitudinal, pink is transverse and green is normal stress.	94
7.8	Variation in the residual stress due to variation in the welding current for the nodes along the line of experiment. Red, green and blue lines illustrate higher bound, $+3\sigma$, mean, μ , and lower bound, -3σ respectively	96
7.9	Comparing FEM and the point-by-point regression for longitudinal component of residual stress along the measurement line. Red line is FEM prediction and green line shows regression approximation. . . .	101
7.10	Uncertainty distribution using regression model for a normal variation of double ellipsoid parameters.	102
7.11	Distribution of residual stress for the node distanced 7.5 millimeter from the middle of the weld path	103
8.1	Strain approach to define the hot cracking susceptibility region. . . .	106
8.2	Ductility curves for some stainless steels [48].	107
8.3	Stress approach to define hot cracking susceptibility region.	108
8.4	Bottom view the mesh employed in this analysis.	109

8.5	Comparing computed thermal results and thermal results from [44].	111
8.6	Stress distribution from [44]. 172 [MPa] stress is applied on side and weld at 30 [mm] from the start point.	112
8.7	Current computed stress distribution. 172 [MPa] stress is applied on sides and weld at 30 [mm] from the start point.	112
8.8	0 and 310 [MPa] stress contours for comparison with Fig. 8.6. 172 [MPa] stress is applied on sides and weld at 30 [mm] from the start point.	113
8.9	0 and 160 [MPa] stress contours for comparison with Fig. 8.6. 172 [MPa] stress is applied on sides and weld in 30 [mm] from star point.	113
8.10	Thermal Profile for 7 powers and speed given in Table 8.3. X is the distance on the weld path from the start point and the peak shows the weld position.	115
8.11	Thermal results for the 13 H-denoted rows, Table 8.6, applied to project 1, Table 8.3.	120
8.12	Thermal results for the 13 H-denoted rows, Table 8.6, applied to project 2, Table 8.3.	120
8.13	Thermal results for the 13 H-denoted rows, Table 8.6, applied to project 3, Table 8.3.	121
8.14	Thermal results for the 13 H-denoted rows, Table 8.6, applied to project 4, Table 8.3.	121
8.15	Thermal results for the 13 L-denoted rows, Table 8.6, applied to project 6, Table 8.3.	122
8.16	Thermal results for the 13 L-denoted rows, Table 8.6, applied to project 7, Table 8.3.	123
8.17	Thermal results for the 13 H1-denoted rows, Table 8.9, applied to project 1, Table 8.3.	126

8.18	Thermal results for the 13 L7-denoted rows, Table 8.10, applied to project 7, Table 8.3.	127
8.19	Variation in transverse component of stress, σ_{xx} , on the weld path for different speed-power-double-ellipsoid settings when traction is 0 N. .	129
8.20	Variation in transverse component of stress, σ_{xx} , on the weld path for different speed-power-double-ellipsoid settings when traction is 1250 N.	129
8.21	Variation in transverse component of stress, σ_{xx} , on the weld path for different speed-power-double-ellipsoid settings when traction is 1875 N.	130
8.22	Variation in transverse component of stress, σ_{xx} , on the weld path for different speed-power-double-ellipsoid settings when traction is 2500 N.	130
8.23	Variation in transverse component of stress, σ_{xx} , on the weld path for different tractions when speed-power-double-ellipsoid setting is row 1 in Table 8.3.	131
8.24	Variation in transverse component of stress, σ_{xx} , on the weld path for different tractions when speed-power-double-ellipsoid setting is row 2 in Table 8.3.	131
8.25	Variation in transverse component of stress, σ_{xx} , on the weld path for different tractions when speed-power-double-ellipsoid setting is row 3 in Table 8.3.	132
8.26	Variation in transverse component of stress, σ_{xx} , on the weld path for different tractions when speed-power-double-ellipsoid setting is row 4 in Table 8.3.	132
8.27	Variation in transverse component of stress, σ_{xx} , on the weld path for different tractions when speed-power-double-ellipsoid setting is row 5 in Table 8.3.	133

8.28	Variation in transverse component of stress, σ_{xx} , on the weld path for different tractions when speed-power-double-ellipsoid setting is row 6 in Table 8.3.	133
8.29	Variation in transverse component of stress, σ_{xx} , on the weld path for different tractions when speed-power-double-ellipsoid setting is row 7 in Table 8.3.	134
8.30	Comparison of the hot crack susceptibility between the best setting, i.e., the fastest welding with zero traction, and the worst setting, i.e., the slowest welding with highest traction. The stress component is σ_{xx} .	135
8.31	Comparison of the hot crack susceptibility for 7 welding speeds when traction is 200 MPa. The stress component is σ_{xx}	136
8.32	Comparison of the hot crack susceptibility for 4 tractions when the welding speed is the fastest. The stress component is σ_{xx}	136
8.33	3D visualization of the Gauss points that satisfy the hot cracking conditions for BTR [1000, 1700] for the project 1-1.	138
8.34	3D visualization of the Gauss points that satisfy the hot cracking conditions for BTR [1000, 1700] for the project 1-4.	138
8.35	3D visualization of the Gauss points that satisfy the hot cracking conditions for BTR [1000, 1700] for the project 4-1.	139
8.36	3D visualization of the Gauss points that satisfy the hot cracking conditions for BTR [1000, 1700] for the project 4-4.	139
8.37	3D visualization of the Gauss points that satisfy the hot cracking conditions for BTR [1000, 1700] for the project 7-1.	140
8.38	3D visualization of the Gauss points that satisfy the hot cracking conditions for BTR [1000, 1700] for the project 7-4.	140

8.39	3D visualization of the Gauss points that satisfy the hot cracking conditions for the worst setting, slowest weld with highest traction. The radius of the spheres shows the transverse stress magnitude.	141
8.40	Top view of a 3D visualization of the Gauss points that satisfy the hot cracking conditions for the worst setting, slowest weld with highest traction. The radius of the spheres shows the transverse stress magnitude.	142
9.1	The cross section and mesh employed in this analysis. Wall thickness is 17.5 [mm].	148
9.2	The notation of sub-passes for the first and second layer. They can have negative sign to show the change in direction.	149
9.3	Comparison between the surrogate approximation constructed based on the algorithm presented in [7], 9.3(a), and the surrogate approximation constructed based on the algorithm developed in this analysis, 9.3(b). The red line shows the exact result from FEM and the other colors show each term in the surrogate approximation.	152
9.4	A snapshot of the thermal result that shows the transient temperature field when the weld is in the middle of the first sub-pass of the first layer. Wall thickness is 17.5 [mm].	154
9.5	A snapshot of the thermal result that shows the transient temperature field when the weld is in the middle of the first sub-pass of the second layer. Wall thickness is 17.5 [mm].	155
9.6	Total displacement and axial stress field when the first layer of weld finished using the sequence (a, b, c) and cooled down to the ambient temperature. Wall thickness is 17.5 [mm].	157

9.7	Total displacement and axial stress field when the first layer of weld finished using the sequence (a, b, c) and cooled down to the ambient temperature. This layer follows the weld shown in Fig. 9.6. Wall thickness is 17.5 [mm].	158
9.8	Position A (blue), B (green), and C (yellow) picked for constructing the surrogate model. Coordinate axes are shown in red, yellow and green for the X, Y and Z respectively	166
9.9	Displacement [mm] results for all settings in the sample space. Directions X, Y, and Z are illustrated in 9.9(a), 9.9(b) and 9.9(c)	167
9.10	Comparison between the displacement [mm] obtained from FEM and surrogate approximation for the sequence (c, a, b). Directions X, Y, and Z are illustrated in 9.10(a), 9.10(b), and 9.10(c).	170
9.11	Comparison between the displacement [mm] obtained from FEM and surrogate approximation for the sequence (b, c, a). Directions X, Y, and Z are illustrated in 9.11(a), 9.11(b), and 9.11(c).	172
9.12	Comparison between the displacement [mm] obtained from FEM and surrogate approximation for the sequence (-a, b, c). Directions X, Y, and Z are illustrated in 9.12(a), 9.12(b), and 9.12(c).	174
9.13	Comparison between the displacement [mm] for the best sequence (c, b, a) 9.13(a) and the worst sequence (-a, -b, -c) 9.13(b). Wall thickness is 17.5 [mm].	176
9.14	Mesh employed for plotting the results. The red line is a reference line.	177
9.15	Radial displacement plotted for the line number 55 which is -50 degree from the reference line. Red shows the best sequence and the green shows the worst sequence.	178

9.16	Radial displacement plotted along the circles number 1, 11, and 14 respectively. Number 1 is the farthest from the weld, number 2 is close to the weld and number 12 is in the middle of the weld. Red shows the best sequence and the green shows the worst sequence.	179
9.17	Hoop residual stress at the end of the process for the best (a) and the worst (b) sequence. These plots are on axial lines number 8, 10,12 , 1, and 20 where the first three are about 45 degrees from the reference line between the tack welds and the last two are on the first and second tack welds. Fig. 9.17(c) compares the line 10 and 20 of both cases. . .	180
9.18	Hoop residual stress at the end of the process for the best (a) and the worst (b) sequence along the circle number 3, 6, 8, 11, and 12 where the number 3 is the farthest and number 12 is the closest to the weld mid line. Fig. 9.18(c) compares hoop residul stress of the line 3 and 12 of the both cases.	181
A.1	Double ellipsoid model showing semi-axes length parameters.	199
B.1	Manipulation of the basic 2D simplex to find the minimum by N-M downhill simplex method.	212
B.2	An example solved by N-M algorithm.	213
B.3	Steps taken by the <i>Compass search</i> algorithm using 4 cores for evaluation of results.	215
B.4	A possible path the algorithm might take in a quadratic objective surface by the coordinate search method.	217
B.5	A possible path the algorithm might take in a quadratic objective surface by the variable coordinate axes search method.	218
B.6	A possible path the algorithm might take in a quadratic objective surface by the steepest descent search method.	219

B.7 A possible path the algorithm might take in a quadratic objective surface by the conjugate gradient search method. 219

Nomenclature

1D	One Dimensional
2D	Two Dimensional
3D	Three Dimensional
ANOVA	Analysis of Variance
BC	Boundary Condition
BTR	Brittleness Temperature Range
CAD	Computer-Aided Design
CPU	Central Processing Unit
CST	Critical Strain Rate
CTE	Coefficient of Thermal Expansion
CWM	Computational Weld Mechanics
DOE	Design of Experiments
FEM	Finite Element Method
GHz	Giga Hertz
GUI	Graphical User Interface

HAZ	Heat Affected Zone
JRM	Joint Rigidity Method
MIG	Metal Inert Gas Welding
MPC	Model Predictive Control
MPI	Message Passing Interface
MSE	Mean Squared Error
N-M	Nelder-Mead
NLP	Non-Linear Programing
NR	Newton Raphson
ODE	Ordinary Differential Equation
R&D	Research and Development
RMSE	Root Mean Squared Error
STL	Stereo Lithography CAD
TBW	Temper Bead Welding
TIG	Tungsten Inert Gas Welding
V&V	Verification and Validation
wrt	with respect to

Chapter 1

Computational Weld Mechanics

Computational weld mechanics (CWM) deals with models, numerical algorithms and software to predict the mechanics of welds and welded structures. CWM applies multi-physics theory of thermal, micro-structure and stress analysis to analyze complex welded structures. The finite element method (FEM) is the overall dominant tool used in CWM. Current CWM models have been shown to be accurate enough to meet the requirements of many industrial projects. They use mature algorithms to predict the welding behaviour in welded structures.

The state of CWM can be understood from John Vitek's question, 2006, "Models are Great, So What?" [1]. This suggests that the necessary level of reliability exists for CWM models and asks why they are not employed regularly in industry.

In a design stage, simulation-based methods have become major tools for product design but are less common for design for manufacturing [2]. Practical CWM is often employed in R&D departments but there is little application of CWM in production design and control. The main limitations comes from the fact that these models are complex because the physics of welding is complex [3]. Therefore, (a) the users require training and experience, (b) the expert-user time required to prepare a weld model to do one analysis can be significant, (c) the CPU time required for one analysis can be long, and (d) actual CWM design problems are not based on solving one instance

of a weld simulation.

Companies could buy software and train staff to setup the simulation but their main concern is the total response time to solve the simulation is too long. The heavy machinery industry, for example, requires an optimal weld sequence to be determined within 30 days and such an optimization requires analyzing a great number of weld projects. Even if one uses many computers to reduce the time to compute the optimization, the limitation of experienced-people time to setup the total number of required projects remains.

Using the right inputs in CWM is another issue from industrial viewpoint. It is always a question what is the right value of an input parameter to be picked in the range of allowable variations to have a realistic analysis. Some parameters are process parameters and some others are given material properties. Most of the industrial measurements done for a property parameter ignore the information about the variations, and as a result there is a risk of using an unrealistic value. If the model is sensitive to either process or property parameters and variations of these parameters are neglected, then the result could be unrealistic. Chemical composition and cut-off-temperature are two examples of such parameters. On the other hand, if the data comes with a range of variations for a set of input parameters, then it is desirable to run analyses to compute the response surface for the configuration of inputs that best matches a reference, i.e., experimental measurement.

Practical CWM problems require a great number of CWM analyses that are not feasible with manual-setup-based frameworks because the implementation time required is too long and human mistakes are very probable. A framework can save an expert-user's time by automating the multiple setups of the CWM analyses required.

1.1 CWM Control and Optimization

A control and optimization application of CWM could be attractive in industry. CWM provides advanced predictive models for optimal decision making in these applications. Coupling the sensor data to CWM has the potential to create the experiments to predict behaviour of weld and adjust the control parameters for real-time weld control. Distortion, for instance, could be an objective to be minimized and parameters of welding process, cool-down times, sequence, and so on could be design parameters for the controller to adjust. Ideally, a framework analyzes possible scenarios using CWM predictive models and picks the optimal solution, e.g., minimum distortion, in a real-time control of a weld.

Model Predictive Control (MPC) [4] applies a model to predict the behaviour of a complex dynamical system, i.e., outputs. The model predictive controller uses the model and current state measurements to give future values of the independent variables that will result in an operation that honours all independent and dependent variable constraints [5].

Mathematically, this solves a nonlinear programming problem (NLP) to minimize a scalar-valued objective function $f(x)$ over all vectors x satisfying a set of constraints. The constraints are in the form of general nonlinear equations and inequalities. Each instance of a control vector $x = (x_1, x_2, \dots, x_n)$ gives one configuration for the variables of the problem.

In a CWM problem with a set of control vectors, each vector x is one design setting and the objective function designates the desired behaviour to be optimized over total design settings. To solve an NLP problem, the objective function must be evaluated for a vector x . In effect, each evaluation is a CWM project using the components of the control vector x , i.e., the CWM parameters. If the NLP problem requires m CWM evaluations, this forms a design of experiment (DOE) matrix with m rows such

that each row corresponds to one evaluation or one CWM project. Each column of the DOE matrix contains the values of the design parameters or n components of the vector x . Depending on the algorithm used in NLP, different DOE matrices or sequence of DOE matrices might be created for evaluation.

The DOE matrices can be quite large in a practical CWM problem. It is a challenging task to implement such m by n DOE matrices because it requires creating m projects, assigning n values to each of m CWM projects, running them and managing the results of m evaluations of the objective function. Considering the complexity of a CWM problem, manual-setup-based implementation of NLP DOE matrices is not feasible in practice.

Everyone makes mistakes. Implementing a large DOE matrix manually without making a mistake in the setup is unlikely. Missing a simple setup task is very likely in multiple manual setups even for expert users. This results, not only in a wrong evaluation or failure in computing, but also requires additional time to find the problem and correct it. Automating the setup for implementing such DOE matrices is a robust way to minimize human mistakes.

Research projects on MPC and NLP for welding processes focus mostly on welding machine parameters such as current and voltage to keep the arc-weld smooth and follow the weld path using small DOE matrices with a quick empirically based evaluation of the objective function.

Control and optimization are mature disciplines. There are a significant number of advanced methods and algorithms to solve a variety of problems. Many of them could be used for solving weld problems. However one might suggest CWM as a prediction model in algorithms but it is hard to implement DOE matrices created in each stage of the algorithms.

Few control and optimization papers have been published in the context of CWM. Michaleris [6] implemented an algorithm for a weld optimization problem using direct

differentiation. This involves computing the derivatives of the governing equations wrt the design parameters. For sufficiently smooth problems, optimization with direct differentiation is expected to be fast and accurate because it utilizes the gradient of the objective function. However, the disadvantage is that the code must be written to compute the gradient of the state equation or residual wrt to control parameters. Michaleris' direct differentiation was the main effort for the derivative-based CWM optimization problem [6].

Derivative-free optimization works with the direct value of the objective function, and therefore the optimization requires the evaluation of results. This creates a DOE matrix or sequence of DOE matrices to be evaluated. Using such DOE matrices is introduced; multiple setups, accumulated expert-people time, possible setup mistakes, and long CPU time in practice. As a result, such optimizations would not be practical without a framework to automate the multiple CWM setups and evaluations.

Voutchkov et al. [7] developed a surrogate model for a weld sequence optimization to minimize distortion in a tail bearing housing by analyzing a DOE matrix of 27 tests chosen from the total space of 46,080 configurations. Tsai et al. [8] developed joint-rigidity-method (JRM) to determine the weld sequence to minimize the distortion in a thin-plate panel structure with 18 welds for Hyundai. This method needs $n(n+1)/2$ analyses, i.e., 171 for Hyundai's thin-plate structure including a sequence of 18 DOE matrices with 18, 17, 16,..., 1 rows.

There are many effective control and optimization algorithms that could be employed in CWM. However, lack of support for a robust and feasible implementation of their DOE matrices limits their practical use.

1.2 Industrial CWM Problems

The fundamental scientific theory behind the weld distortion and residual stress mitigation techniques and related design space is well understood. In 1950s, [9], Vinokurov [10], and Brust [11] discussed a variety of techniques to minimize distortion and residual stress in welded structures. Even so, the capability to solve realistic CWM problems is only reaching maturity now.

Exploring the design space to determine an optimum configuration can require hundreds of CWM projects to be evaluated. Even if one finds an optimal configuration for a problem, it is not necessarily the optimum for another problem and every new problem needs its own collection of CWM projects to be evaluated.

Another typical industrial question is the weld sequence optimization. The answer is to explore a discontinuous design space with isolated points. Theoretically, many algorithms are available for solving such combinatorial optimization problems. One can either evaluate the full or reduced space, but the design space can be very large, i.e., $2^n \times n!$ for n weld passes, or CWM complexity limits the number of evaluations for realistic industrial problems. Volvo [7], Hyundai [8], and a few other companies have tried to develop CWM-based methods to find the best sequence of welds or weld passes but the methods are not practical for routine design optimization because the response time from design to production is beyond the allowed time.

A robust design is always sought in the design stage and it requires a sensitivity analysis wrt design parameters. The welding process is one of the main concerns in a robust design and the designer needs to know how the final part is affected by variations in welding parameters. Michaleris [6] did a CWM sensitivity analysis based on direct differentiation. Unless the design space is small, sensitivity algorithms using analysis of variance (ANOVA) DOE matrices have similar limitations as discussed for practical use.

Geometric permutation of parts under welding is now studied at the University of Lulea using CWM. Parts are not positioned exactly in a production line and each sub-part is a welded structure has a slight rigid body movement that affects the final tolerance of the welded structure due to tolerance stack-up. To make it robust, the analysis requires a sequence of DOE matrices to be implemented for different position configurations. There exist many advantageous fields of using CWM in industrial problems but little actual usage. One important reason could be the limitations of manual-based implementations to set up multiple CWM projects in a DOE matrix.

In routine engineering in industry, the design group designs a welded structure for a certain performance assuming that welds can be made, then welding engineers develop weld procedures for the design including welding process parameters, number of passes, and so on. After that, manufacturing engineers determine the fixtures, manufacturing design and setups to have the welding done. The manufacturing design can have a significant influence on the initial design such that a feedback is sent to the design group for changes, improvement, or optimization. This design loop of optimization is too long and expensive in the range of several years' time and costs because the design group has very limited information on how good the design is. An automated implementation of CWM to support multiple setups for optimization in an affordable time and cost as a tool to support designer-driven optimization could improve the design loop.

Chapter 2

Control and Optimization Problems

From a control point of view, we have a control problem with a large time-dependent vector Q that describes the state of the underlying system at the time t . The state is governed by a state equation written as a form in Eq. 2.1 where u is a control vector and F is a mapping from $R^m \rightarrow R^n$. R^m forms the control space and R^n is the response space [12].

$$Q_t = F(Q_{t-1}, u_t) \quad (2.1)$$

We also define a scalar-valued function ϕ called an objective function that is to be minimized or maximized. ϕ is a function of state, Q , and control vector u . The state, Q , is a function of the control vector u and therefore the function ϕ really depends only on u (See Eq. 2.2).

$$\phi(Q, u) = \phi(Q(u), u) = \phi(u) \quad (2.2)$$

The question that all optimization methods try to answer is how the objective function, ϕ , changes wrt changes in the control vector u . Derivative-based optimization algorithms try to compute or approximate the derivative of ϕ wrt u , $\partial\phi/\partial u$, and derivative-free algorithms require the value of ϕ for a given u .

2.1 Objective Function

The objective function ϕ must be able to measure how well the user's goals are met. There is no general form of such a function and the specific form of such function comes from the nature of the problem and creativity of user. However a p-norm could be a good function for many cases [13].

Stam et al. [14] defined Eq. 2.3 as an objective function that could be used in a control problem over time. A set of key-frames vectors Q^* for the state as the user's reference is defined. There is a scalar function ϕ for each time step that measures how closely the result is to the key-frames and also penalizes the system for using too much control vector u . They have a weight vector, W , to give higher weight to certain regions of space. The penalty terms can be added as a smoothness coefficient, α , to the correction in state Q due to the change in control vector u in each time step.

$$\phi(u) = \frac{1}{p} \sum_{t=0}^n (\| W_t (\gamma(Q_t) - \gamma(Q^*)) \|^p + \alpha \| Q_t - Q_{t-1} \|^p) \quad (2.3)$$

Here γ is a pre-processing function or procedure applied on the state for each time step. It is for user to pick a function and a p-norm could be a good function. L_p norm is defined in Eq. 2.4 and $p = 1, 2, \dots$, are frequently used [13].

$$\| x \|_p = (|x_1|^p + |x_2|^p + |x_3|^p + \dots + |x_n|^p)^{1/p} \quad (2.4)$$

2.2 Optimization Methods

Historically, most approaches to optimization use a truncated Taylor's series expansion to a local-linearized approximation of the objective function in a sense that uses the low order derivatives of the objective function for approximation and ignores the

higher order derivatives. In fact, one can classify most methods for numerical optimization according to how many terms of the expansion are exploited [15]. Newton's method, for instance, is a second-order method based on the second-order Taylor polynomial approximation and therefore the first and second derivatives should be available. Steepest descent, another instance, works with the first derivatives and local-linearized approximation based on the first-order Taylor polynomial and it is classified as a first-order method. In this taxonomy, zero-order methods rely exclusively on values of the objective function and do not require derivative information. They are direct-search methods, which indeed are often called zero-order/derivative-free methods in the engineering optimization community [15], [16].

The algorithms using the derivative are somewhat more powerful than those using only the function values [17]. However they are limited in that firstly both the objective function and the design parameters have to be continuous and secondly if the objective function contains multiple extremum, then a gradient method converges to the local extrema in its basin of attraction with little chance to find the global extremum. If such cases, then derivative-free approaches that work with direct value of the objective function can solve the problem [18].

Appendix B reviews optimization methods. The focus of this research is primarily on derivative-free methods because they are more attractive in a framework that automates the implementation and evaluation of DOE matrices created by these methods.

2.3 Parametric Design Space

A parametric design application can be characterised as a mapping from a multiple-dimensional space $\langle u, \delta, C, R, P, \phi \rangle$ to a set of solution designs, $\{D_1, D_2, \dots, D_n\}$, where in CWM: u could be vector of CWM parameters, δ could be step size for each

parameter, C could be a set of constraints such as bounds on each parameter, R could be requirements such as minimum step size, P could be preferences in a set of Pareto front, i.e., the set of all global optima (Pareto-points) or an approximation in the context of multi-objective function optimization, and ϕ could be the objective function to be minimized. Each point in the design space is one design state/configuration that corresponds to one CWM project setup. Choosing a set of sample points in the parametric design space defines a DOE matrix such that each point is one row in the DOE matrix and each column corresponds to one parameter.

The parametric space for our computer model for CWM has about 300 parameters. Most optimization and control applications of CWM are defined on a small sub-space of the parametric space often with dimension less than 10 parameters. The CWM parameters can be categorized as below.

- **Weld Process**, e.g., process type, weld power, traveling speed, double ellipsoid shape, and so on.
- **Weld Joint**, e.g., weld path, weld start/end time, start/end position, delay times, and so on.
- **Weld Sequencing**, e.g., number of sub-passes/weld path, different sequencing patterns, sequential/simultaneous patterns, inter-pass temperature, and so on.
- **Fixtures and Boundary Conditions (BCs)**, e.g., Dirichlet/Neumann BCs, clamping position, apply/release time, convection coefficient, contact parameters, and so on.
- **Chemical Composition**, e.g., carbon, alloying components, uncertainty range, and so on.
- **Material Properties**, e.g., specific heat, thermal conductivity, initial grain

size, initial hardness, Young modulus, Poisson's ratio, yield stress, hardening modulus, and so on.

- **Thermal, Microstructure, and Stress Simulation Parameters**, e.g., solver type, heat source model used, number of Newton-Raphson (NR) iterations, convergence criteria, and so on.
- **Meshing and Parts**, e.g., mesh type, level and type of refinement, coarse-fine perturbation, rigid body movement, contacts, and so on.
- **Initial State**, e.g., data flow from other projects, re-start time step, result mapping, and so on.

These parameters can vary in time or space of welding and cool-down after welding. Increasing the number of parameters enlarges the parametric design space. A parametric design with about 5 to 10 parameters is a quite large space to explore. Considering that a CWM model can have roughly 300 parameters, unless a small subset of parameters is selected, exploring a parametric design would not be feasible. For example, to compute Taguchi's sensitivities [19] for 15 design parameters with 3 levels, L_{36} requires 36 analyses to screen the parameters. Selecting 8 parameters with with 3 levels for a full factorial analysis requires $3^8 = 6,351$ analyses to compute a local approximation to the response surface, i.e., approximation to the state manifold or behavior of the system. A fractional factorial analysis still requires a large number of analyses. To compute a response surface requires many more design points. Using an automated framework and depending on the problem's characteristics, exploring a parametric design with less than 10 parameters could be feasible.

Parametric design analyses must solve tens or hundreds of design points to explore or map the associated design space specified by a parametric design DOE matrix to find optimal designs. In a parametric design, algorithms that use DOE matrices for more efficient searching are preferred.

Such DOE matrices takes advantage of the fact that multiple trial solutions can be obtained simultaneously in contrast to sequential algorithms that do search based on solving one-problem at a time. In a framework for exploring a parametric design space using DOE matrices, the number of cores allocated to the problem affects the number of rows in the DOE matrix. To use the available cores efficiently, the total number of projects, (rows) in a DOE matrix should be divisible by the number of cores that are to be utilized.

If the number of cores is twice the number of design parameters, the DOE matrix can afford a full factorial perturbation of $\pm\delta$, step size, at any trial point in the same computing time as a single analysis on a single core. If the number of cores is half or quarter of the number of parameters, a fractional factorial DOE matrix can be analyzed in the same time as a single analysis.

If the algorithm is based on random sampling, a larger number of random samples can be analyzed simultaneously.

If the algorithm is based on the random crossings in the population, e.g., a genetic algorithm, the number of offspring can be increased.

If the algorithm constructs an approximation model, an estimator can be designed to obtain a better approximation from more sample points.

If the problem involves finding a global extremum and there is a chance of being trapped in a local extremum, the algorithm can use a feasible coarse grid over the design space and sample a larger region of the design space simultaneously.

Generally, most of optimization algorithms become more efficient by using parametric design DOE matrix on multi-cores. This motivates designers to exercise a wider range of optimization algorithms in CWM since they become more affordable for practical CWM.

2.4 CWM Optimizer Framework

The concept of the optimizer framework for CWM can be summarized with the following modules:

1. Start with a base project.

The base project is a reference project that an optimizer employs to apply a DOE matrix. This base project can be a project setup in a simulation code. The setting from this project is also used as initial trial point in the design space unless the user defines a different one. This can be a module called, for instance, *Base Project*.

2. User defines objective function(s).

The objective function is a scalar-valued function that is to be optimized. This scalar number is the value of the solution for each design point and determines the merit of the solution. This can be a module called *Objective-Function Evaluation*. Some examples could be; (a) predefined terms including L_1 , L_2 , L_∞ norms, integral, sum, MSE (mean squared error), standard deviation and so on for a set of results such as nodal displacement, residual stress, hardness and so on: (b) sequence of actions introduced by user and recorded as a macro from a GUI to evaluate a certain quantity of interest.

3. Choose the sampling method to create and apply a DOE matrix.

Design space contains many design states that one of which shows a design configuration. Either it is affordable to test all states in the space or one must use a reduced-space method with a smaller sample space. An $m \times n$ DOE matrix must be created to assess m design states with n parameters of the design configurations. Some example DOE matrices [20] could be full factorial, reduced factorial, D-optimal, Taguchi, Monte Carlo, Box-Benken, hypercube,

adaptive, discretized design space, optimization-algorithm defined or user defined. When this DOE matrix is defined, the machine creates the m number of projects needed and assigns the n design parameters from the DOE matrix to each project. This can be a module called *Applying DOE Matrix*.

4. Choose an optimization algorithm for finding the extremum.

To solve a single/multi objective function unconstrained/constrained problem, there is no one-size-fits-all optimization algorithm. The characteristics of the problem define the preferred algorithm for a given problem. Therefore the user must choose the optimization algorithm. Some example algorithms could be downhill simplex, direct search, surrogate/meta model, least-square direct search, full/reduced factorial DOE matrix, radial basis function, genetic, simulated annealing, Kriging, neural network, and others. Parametric design DOE matrix is used in this stage to create an efficient DOE matrix. Examples for direct search, surrogate/meta model, least-square direct search and full/reduced factorial DOE matrices are presented in Appendix B. This can be a module called *Optimization Algorithm*.

5. Allocation manager to send the jobs to available resources.

After applying DOE matrix to m projects, the allocation manager organizes the machine, cores and memory to obtain the assessment results as quickly and cheaply as possible based on the available recourses. A parametric design DOE matrix requires information on available machines/cores and this manager provides this information. Clustering and parallel computing, for instance, could be employed. This can be a module called *Allocate Projects to Cores*.

6. Decision support to choose the optimum.

If there is more than one optimum and/or the optimal solutions are in a Pareto front, the user must be involved in the decision-making algorithm, and the

framework should generate the information needed for the user to make decisions. This can be a module called *Decision Support*.

7. Post-processing any user-requested output from the results. This can be a module called *Post-Processor*.

Fig. 2.1 shows how these modules interact with each other in the framework. In the proposed framework, the tasks for user(s) is to setup one base project, design DOE matrix, define an objective function, and make decisions. Machine starts with the base project and creates m sub-projects linked to the base project as a master project. Given n columns of DOE matrix are applied to the sub-projects. The sub-projects are allocated to available cores for evaluations. When the evaluation is finished, the objective function is looped over the sub-projects and the outputs are collected for the user to make a decision. These simulation results might be removed to save disk space. If the decision is fine, it stops otherwise another DOE matrix will be designed for the next iteration.

So far, the modules *Base Project* and *DOE* are currently supported in VrSuite [21]. Module *Objective-Function Evaluation* is partially supported in VrSuite. Module *Optimization Algorithm* is linked to the other modules but the decision-making is still user-based and it is not automated in VrSuite. *Allocation Manager* is developed to work on one machine with multi-cores and an advanced one, i.e., message-passing-interface (MPI), is under development. Post-processing is supported in VrSuite.

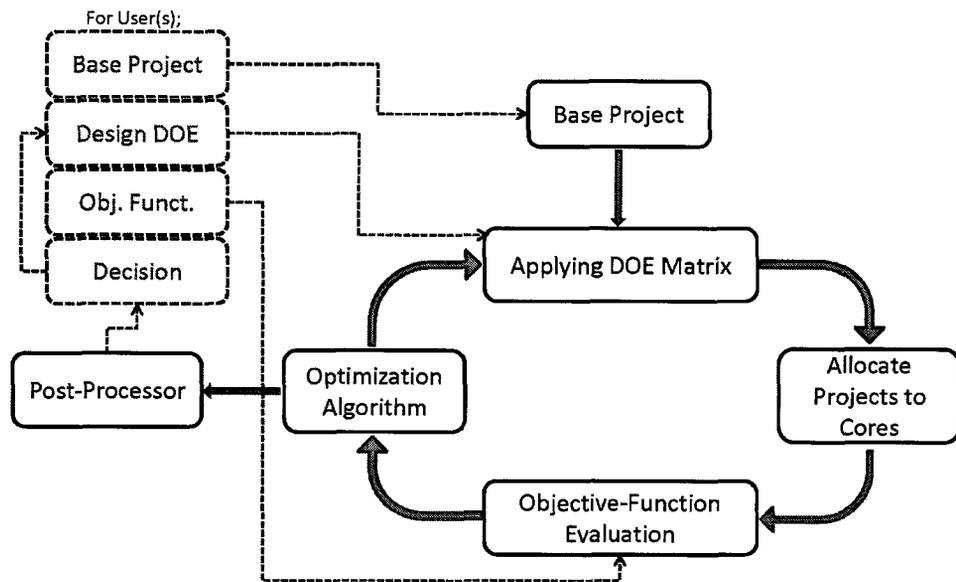


Figure 2.1: Suggested optimizer framework

Chapter 3

CWM Design Space Exploration Using DOE Matrices

A CWM multi-project optimizer framework was discussed and a general concept was proposed in the previous chapter. To implement and develop this framework, the strategy is to provide and improve the support required for exploring the design space of real CWM problems using different algorithms and DOE matrices to test and explore the aspects for the CWM multi-project optimizer framework and the modules development.

3.1 Validation of CWM

Validation of a Computational Weld Mechanics (CWM) code for a particular welding application requires an estimate of the difference between experimentally measured parameters and parameters computed by a computational model. This requires estimates of the uncertainty in both the experimental data and the computational data and this, in turn, requires careful design of both the experiment and the CWM model. ASME V &V standard [22] gives a guide line, Even so, the experimental measurements for welding is usually one (or in some case two) trial measurement with no

information about uncertainty quantification and correlation between the parameters. Having a valid experimental results, then it is for the CWM to minimize the error in the control vector. The author's opinion is that the main limitation is the lack of valid experimental data. Chapter 4 shows the current research group's experience in performing validation tests for a Computational Weld Mechanics (CWM) code. It suggests that the greatest source of error is in specifying the control vector, i.e., the parameters that characterize the weld experiment. Once errors in the control vector have been reduced sufficiently, then errors in material composition, temperature dependent material properties, and evolution of microstructure become important.

3.2 Design Space Discretization

A continuous problem with a quite large range of possible variation in the design parameters could be discretized by a certain step size for each design parameter and use a full or reduced set of discretized points in a DOE matrix to give a well-covered map from the design parameters to a response surface. This approach could be effective if the response surface does not fluctuate too much, the number of design parameters is small or there are many available machines/cores to run a great number of analyses at the same time. This also could work effectively if the design parameters can be categorized with no significant correlation between the categories. Another usage is to test an initial coarse grid over a large domain in search for a global extremum. This method might need to continue using a sequence of adaptive refined grid DOE matrices. Chapter 5 describes an analysis that solves an optimization problem using this strategy to minimize distortion and show the sensitivity of the distortion of an edge weld on a 152 x 1220 x 12.5 [mm] bar of Aluminum 5052-H32 wrt the parameters of clamping design space. In the first part, the design space is discretized by a grid of 5 by 9 resulting in 45 points in the discrete design space. A

full factorial DOE matrix of 45 points is used to give a fully-covered map from the design parameters to the response surface. In the second part, a grid of 44 points used to be divisible by 4 available cores and save CPU time.

3.3 Regular and Least-Square Direct-Search

In a continuous response surface case, if the initial DOE matrix cannot find the optimum, there are direct-search algorithms to learn from DOE matrices that have been evaluated and find a possible path toward an optimum. It, then, creates another DOE matrix for the next move. The algorithm repeats the learning from the previous results to follow the path until it reaches the minimum or some imposed limits. Different direct-search algorithms are available. A regular one moves to a new optimum chosen from the result of the last DOE matrix and forms a new DOE matrix for the next iteration. Least-square direct-search uses the least-square approximation to estimate the response surface and moves to the optimum point using the approximation model to guess the location and form a new DOE matrix for the next iteration. Chapter 6 describes an analysis that presents both regular and least-square direct-search methods for solving a constrained optimization problem. This chapter concludes that the least-square direct-search algorithm follows the path to the minimum more efficiently in the neighbourhood of a smooth basin. However the least-square approximation is not expected to work well when the response surface is not a smooth basin. For example, if the response surface is very wavy or rough. Either regular or least-square direct-search generates a sequence of DOE matrices with different size using different number of cores.

3.4 Regression/Response Surface Analysis

A reliability-based design needs to know how the uncertainty in design parameters changes the results. Monte Carlo is a robust algorithm to use in such cases. Increasing the number of design parameters raises the number of points in the sample space and therefore the number of rows in DOE matrix for the Monte Carlo analysis. Chapter 7 describes an analysis that uses the Monte Carlo method to find the uncertainty distribution from weld parameters. The first section deals with one parameter, i.e., welding current, and a normal distribution generates a DOE matrix of size 30 for different welding currents utilizing 4 cores. Using an accurate model, this DOE matrix gives the distribution of residual stress, i.e., objective function, wrt the variation in welding current. The second section has four parameters of arc-weld shape and a DOE matrix of 81 evaluations are used to construct a regression response surface as an estimator. This response surface is used for a Monte Carlo analysis with 10,000 random sample points from possible values of the parameters to observe the variation in the residual stress in the space of these parameters.

3.5 Control and Sensitivity Analysis

In a continuous response surface case, some parameters in the design space might require to meet an imposed constraint. For example, arc-weld shape needs to be adjusted when the weld procedure parameters have changed. This is a control problem to find a configuration of four arc-shape parameters to keep the right shape. This control problem needs to explore the design space of the four parameters to find the best configuration of the parameters. Chapter 8 describes an analysis that presents such design space exploration to control the arc-weld shape as a part of a sensitivity analysis. The objective is to find sensitivity of hot cracking in a $50 \times 25 \times 0.25$ mm

316 stainless steel welded by TIG process with a constant transverse force applied on the side surfaces, wrt weld procedure parameters and transverse traction. Sensitivity analysis has its own DOE matrix designed for different weld procedures and tractions. The weld procedures in the sensitivity DOE matrix uses an internal control problem to adjust the arc-weld shape by 9 DOE matrices designed to explore the control design space. This project employed a post processor that uses the data flow from thermal and stress analysis to decide if an FEM Gauss point is susceptible for hot cracking. This shows a capability of using a user-written post processor in the multi-project CWM design space exploration.

3.6 Surrogate/Meta Model

A combinatorial optimization problem, e.g., weld sequence optimization, has a discontinuous design space with a set of isolated points. The simplest approach for a combinatorial problem is to run all possible combinations to find the optimum. When one cannot afford to generate the results for all points in the design space with an accurate model, reduced space methods can be used. These methods work on a sub-space of the design space instead of the total space. Surrogate models are one kind of a reduced space method. They construct an estimator from the information acquired from a sample space, to approximate the results over the complement space of the sample space. Compared to an accurate model, the surrogate models should be significantly cheaper so that it would be feasible to use them over tens of thousands points in a design space. Although surrogate approaches can be efficient, a concern is that users have to invent a recipe for each specific case. Chapter 9 describes an analysis in which a surrogate model is demonstrated to minimize the distortion in a girth weld of a pipe welded with six sub-passes. This combinatorial optimization starts with a DOE matrix with 6 analyses for the first layer to find the minimum

distortion at the end of the first layer. Thereafter, it initializes the stress state of this minimum as the starting point for the analysis of the second layer. It uses a DOE matrix with 14 analyses chosen from a total of 48 combinations to construct a surrogate model that approximates the distortion for the remaining combinations. A full factorial DOE matrix including all combinations is tested to check the accuracy of the surrogate model over all points in the combinatorial space.

3.7 Comparison of DOE Matrices Used in the Projects

Chapter 10 compares the analyses presented in chapters 5, 6, 7, 8 and 9 in terms of DOE matrices used and CPU time. It demonstrates that CWM design space exploration using DOE matrices is now practical for control and optimizing many decisions in the design of welded structures.

3.8 Additional Projects

The three projects discussed below, have employed the multi-project framework using the DOE matrix for the analysis. They are not directly related to the development of the framework and they do use the concept for part of their respective analyses. Therefore no details on the specifics of each analysis are included in this thesis for these projects. The reason for including them here is to show the application of CWM in the context of microstructural analysis and the conclusions obtained.

3.8.1 Optimizing Temper Bead Welding by Computational Weld Mechanics and Design of Experiment Matrix

This analysis describes a computational weld mechanics model to compute the transient temperature and transient microstructure evolution in temper bead welds. Combining this analysis with a Design of Experiment (DOE) methodology, can optimize the design of temper bead weld procedures. Because hardness can be measured easily, quickly and economically, there is a particular focus on hardness as the most important criterion required in practice. Hardness is a function of carbon equivalent, austenite grain size and cooling rate between 1000 to 800 [$^{\circ}K$]. This motivates a DOE matrix for perturbation analyses of two main parameters; the power per unit length, welding current, and welding speed. The result shows the upper and lower bound expected for hardness due to the variations of each parameter.

3.8.2 Power per Unit Length of Weld Does not Characterize a Weld

A number of publications have characterized low-alloy steel welds by specifying the power per unit length and this study shows that the power per unit length of weld is not a unique descriptor of a weld. This analysis uses a 6 by 2 DOE matrix for 6 different welding speeds and currents with the same weld power per unit length of weld. This study uses a 3D numerical model to compute the transient temperature and evolution of the microstructure for this DOE matrix. The results show that the welds have large differences in their transient temperature field and the fields for phase fractions of gamma, martensite, bainite and ferrite-pearlite phases. Also the gamma phase grain size and hardness profiles are quite different.

3.8.3 Microstructure and Hardness Modelling: Effect of Multiple Bead Deposition in Temper Bead Welding Technique

This study investigates whether the sequence of the filler metal deposition can temper the microstructure properties and reduce the hardness in the Heat Affected Zone (HAZ) and parent metal. Four weld beads are deposited in laboratory on a mild steel plate, with each being tempered by an overlapping bead using the Temper Bead Welding (TBW) technique. Numerical simulation is performed using FEA software VrWeld [21] to evaluate the microstructure, hardness and size of the HAZ. It used a validation process with a sequence of DOE matrices to calibrate the parameters wrt the experimental measurements. The study concluded that there are significant improvements in the microstructure, hardness and size of the HAZ as a result of the deposition of the overlapping, tempering beads.

Chapter 4

Challenges in Validation of CWM

Software to Compute Residual Stress and Distortion in Welds

In this chapter, thesis-related experience gathered while performing validation tests for a Computational Weld Mechanics (CWM) code is summarized. In particular, a validation test for a welding application that compares measured and computed transient temperatures, displacements and strains is described in detail. It clearly demonstrates that a computer model can accurately predict this data. Challenges on both the experimental side and computational side are discussed but the greatest challenge is the limited availability of experimental data that includes a measure of the uncertainty in the experimental data.

4.1 Analysis Background

Arc welding is a complex manufacturing process in which an electric arc generates heat in an area approximately $1 \text{ [cm}^2\text{]}$ while travelling at speeds of approximately 1 to 5 [mm/s] to form a pool of liquid metal with a volume usually slightly less than $1 \text{ [cm}^3\text{]}$. As the liquid metal on the trailing edge of the weld pool solidifies, a weld

joint or weld overlay is created. The transient temperature field drives metallurgical phase changes, thermal expansion-contraction, and changes in temperature dependent properties of the materials. The microstructure evolves as it is heated and cooled with phase changes, annealing, recrystallization and grain growth. The strain due to thermal expansion-contraction, phase changes and restraint of the structure and fixtures generates stress and plastic deformation. The plastic deformation leads to residual stress and deformation.

The physics and mechanics of welding can be separated into the physics of the arc and weld pool on the one hand and the physics of the solid on the other hand. The physics of the arc and weld pool primarily involve magneto-hydrodynamics and fluid flow with length scales less than 1 [mm] and time scales less than 0.1 second. The physics of the solid primarily involve solid mechanics with length scales greater than 1 [mm] and time scales greater than 1 second. The coupling between the physics and mechanics of the weld pool and the physics and mechanics of the solid is almost entirely through the thermal analysis in the solid. The heating effect of the arc and weld pool can be modelled by parameterizing a power density distribution function with net weld power and the weld pool shape, size and position as a known function of time [23]. These weld pool shape and size parameters are usually estimated from macro-graphs of cross-sections of the weld joint.

A computer model to predict the behaviour of this complex process can save time, money, and explore “what if” scenario’s that are difficult to assess by experiment. The arguably best models of the welding process, solve the conservation of energy, mass and momentum, i.e., a system of three coupled partial differential equations, with realistic 3D geometry of the structure and the welding process and the most realistic constitutive equations and weld process models available. Because the welding process is transient, non-linear, and usually involves complex geometry, the best model are based on 3D transient non-linear Finite Element Analysis (FEM). Because

such models require complex numerical algorithms and sophisticated software, the first models did not emerge until the 1980s. Current models often have tens to hundreds of thousands of 8-node brick elements and are often solved for thousands of time steps. Reference [24] is good reference on the state of the art of CWM as it emerged in 1985. References [25] and [26] discuss progress in CWM in 1995 and 2008 respectively that has contributed to the development of CWM. VrWeld is largely based on research led by Dr. John Goldak's research in computational weld mechanics.

The essence of the validation process is the comparison of the output or prediction from a computational model with the output or measurements from a set of experiments. To make the discussion definite, a comparison in some detail of the measured data from an experiment described by Masubuchi [27] and the values predicted by VrWeld [21] is considered. In this case, variables to be compared are the transient temperatures measured by four thermocouples, the transient longitudinal strains measured by four strain gauges, and the transient displacement measured by one displacement gauge.

4.2 Computational Weld Mechanics as a Control Problem

Computational Weld Mechanics (CWM) can be viewed as a control problem that has a state vector and a control vector. The state vector represents the geometry or configuration of the structure, the transient temperature, displacement, strain and stress fields, and the evolution of microstructure. Associated with the state vector, are the constraint equations that conserve mass, momentum and energy, and the constitutive equations associated with material properties which are functions of temperature and microstructure and deformation history. The control vector has

parameters such as weld power, weld speed, the distribution of power density near the arc, and the position of the arc. The control vector is a function of time.

In validating a CWM software program, it is useful to distinguish between errors in the state vector and errors in the control vector. If one changes the control vector, then one models a different weld.

The input data to VrWeld are:

1. The 3D geometry of all parts and fixtures in the structure to be welded are usually imported as Stereolithographic (STL) files created in a CAD system.
2. The composition, temperature dependent thermal and mechanical material properties for each alloy type for each part in the structure including filler metal of welds must be specified.
3. The weld process parameters including weld current, weld voltage, and power density distribution in the weld pool are needed. For a MIG welding process the filler metal wire size and speed are needed. The start time and start position of each weld pass, and the arc-travel speed along the weld path are needed. The delay time between the end of one weld pass and the start of the next weld pass is important.
4. If fixtures are applied or released, the appropriate time dependent boundary conditions must be defined to model the time of applying each fixture.
5. A cool-down time after the welds have been completed is chosen to allow the structure to cool down to room temperature.
6. The initial state of the temperature, microstructure, and strain or stress is needed. Often, the initial state of the temperature is the ambient temperature; the initial state of strain and stress are often unknown and assumed to be zero tensor fields but any known initial state can be applied.

7. If sensors such as thermocouples, strain gauges, displacement transducers, residual stress sampling volumes are of interest, then the position, geometry, and the times of interest of each transducer should be specified as virtual sensors.

The output data from VrWeld include the following fields:

1. Transient temperature field.
2. Transient displacement field.
3. Transient elastic, thermal, plastic, and total strain vector or strain tensor field.
4. Transient stress, deviatoric tensor field, and principal stress fields as scalars or vectors.

4.2.1 Guidelines for Verification and Validation for Use in an Application

The preface of the ASME standard [22] states:

This document provides general guidelines for implementing V&V of computational models for complex systems in solid mechanics. The guidance is based on the following key principles:

1. *Verification (addressing programming errors and estimating the numerical errors) must precede validation (assessing a model's predictive capability by comparing calculations with experiment).*
2. *The need for validation and the associated accuracy requirements for computational model predictions are based on the intended use of the model and should be established as part of the V&V activities.*

3. *Validation of a complex system should be pursued in a hierarchical fashion from the component to the system level.*
4. *Validation is specific to a particular computational model for a particular intended use.*
5. *Simulation results and experimental data must have an assessment of uncertainty to be meaningful.*

Although the state of the art of V&V does not yet lend itself to writing a step-by-step performance code/standard, this guide provides the computational solid mechanics (CSM) community with a common language and conceptual framework to enable managers and practitioners of V&V to better assess and enhance the credibility of CSM models. Implementation of a range of V&V activities is discussed, including model development for complex systems, verification of numerical solutions for governing equations, attributes of validation experiments, accuracy requirements, and quantification of uncertainties. Remaining issues for further development of a V&V protocol are identified. An over view of the process of verification and validation is shown in Fig. 4.1.

4.2.2 Verification of VrWeld

Verification of VrWeld software as described in Fig. 4.1, is done by comparing results computed by VrWeld with relevant mathematical problems for which the exact solution is known. Occasionally when mathematical problems with exact known solutions are not available, numerical solutions that are considered to be highly accurate, e.g., highly refined meshes, are sometimes used. This is done for all mathematical equations in the model. Goldak Technologies Inc. (GTI) does verification by creating ‘Test Suites’ that are collections of problems for which the exact solution is known

Fig. 4 V&V Activities and Products

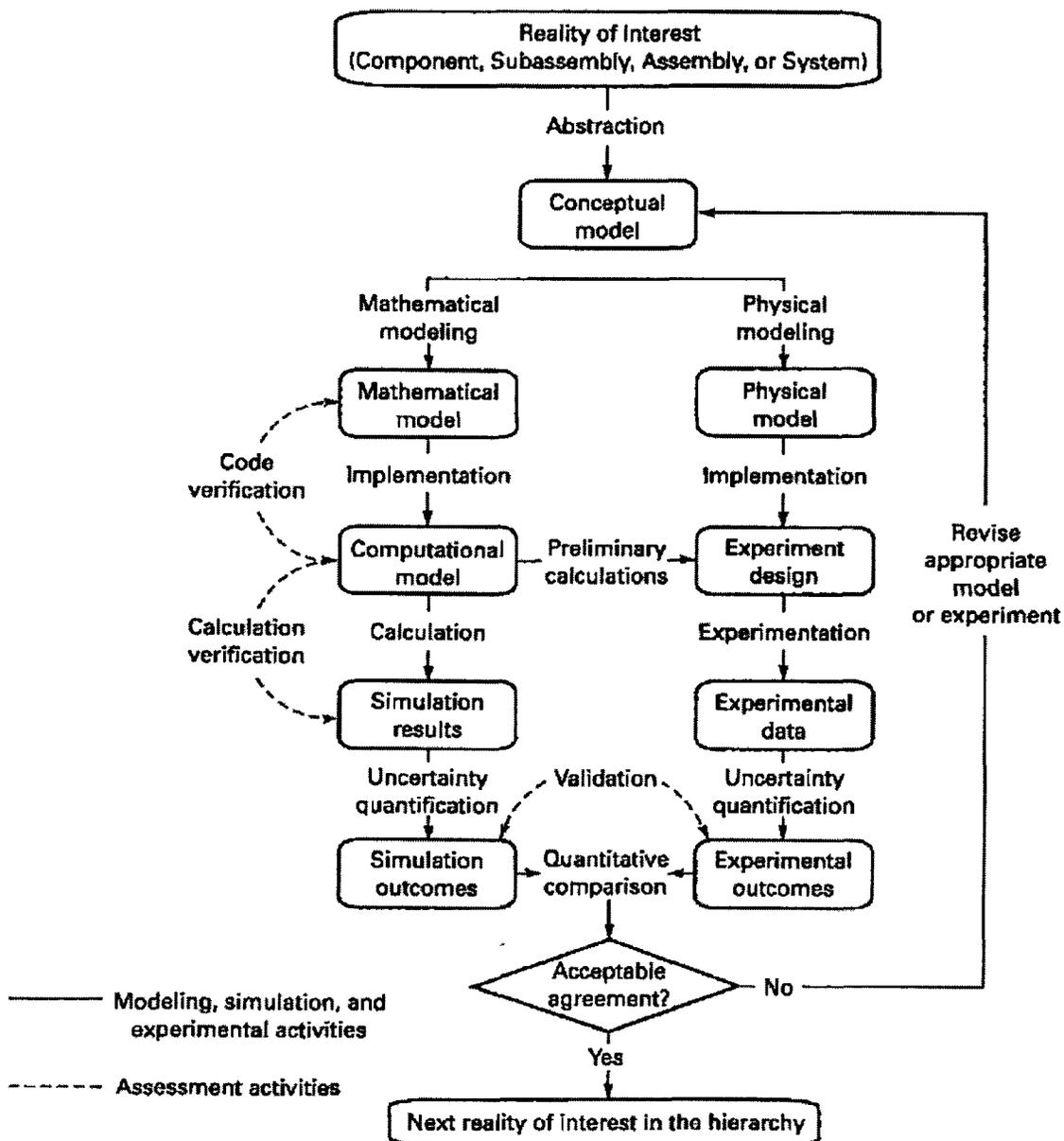


Figure 4.1: This figure taken from [22] provides an over-view of the verification and validation process. The left side is devoted to the development of the model. The right side is devoted to the development of experimental data.

for a particular solver such as the solver for the transient heat equation or the solver for transient visco-elastic-plastic stress analysis. When these test problems are run, the sum of squares error in the solution of each problem is reported.

4.2.3 Validation of VrWeld

Examples of validation are given in the references [28], [29], and [26]. In [28], two examples are an overlay weld analysis with 27 weld passes in two layers covering a 100×100 [mm] block in a 400×400 [mm] square plate 30 [mm] thick and an overlay weld analysis with 51 weld passes in two layers covering a 200×200 [mm] block in a 400×400 [mm] square plate, 30 [mm] thick. The experimental data was gathered from 6 thermocouples and 22 strain gauges. In addition, distortion was measured by a coordinate measuring machine and compared to computed distortion or displacement. The agreement between measured and computed data values was considered to be remarkably good. In [26], the transient displacement field was measured with a high precision stereo camera and compared with the transient displacement computed with VrWeld. Again the agreement was considered to be exceptionally good. In [29], the residual stress measured by neutron diffraction and x-ray synchrotron diffraction was compared with the residual stress computed with VrWeld. Again the agreement was considered good.

In all of these validation tests except [26], only one experiment has been done. It is desirable to conduct multiple experiments and use statistics to assess the uncertainty in the experimental data.

In Computational Weld Mechanics (CWM), the geometry of the filler metal added and the weld pool are important. The sensitivity of the computed solutions to errors in these inputs can be computed by solving the problem with variations in the inputs.

In all validation analyses performed by the current research group to date, the initial state of the structure being welded has been assumed to be zero initial stress

and the geometry was the exact geometry of the original design. If estimates of the initial stress state or deviations from the original geometry were available, then “what if” analyses could be run to estimate their effect.

In addition to errors in the inputs, truncation errors can arise because the mesh size or spatial discretization is finite, and the time step size is finite. These truncation errors can be estimated by running analyses with a finer mesh or shorter time steps. The analyses have been done with linear FEM elements, i.e., 8-node bricks and 6-node prisms. The analyses could be run with either a finer mesh or with a mesh with higher order elements such as quadratic brick elements to obtain an estimate of the truncation error due to finite size of elements. Errors can also be bounded by using duality, e.g., the Prager-Synge Hyper-Circle theorem [30].

After running a first analyses, it can also be useful to identify and form a sub-project on a small sub-domain that has an above average error norm. Then mesh the sub-domain with a finer mesh and solve this small sub-domain with time dependent boundary conditions obtained from a previous analysis with a coarser mesh and possibly coarser time steps. If the sub-domain is smaller than the global domain, it is usually feasible to solve it with a much finer mesh and possibly much shorter time steps than the larger global domain.

4.3 Validation Process

The validation strategy that the current research group usually uses, is summarized below. This assumes that welds have been made with several thermocouples, strain gauges and possible displacement transducers. It could be argued that the validation process as described below should be called a calibration process because the computational analysis has access to the experimental data. However, the desire is to develop a validation process that is accepted by people doing the welding, i.e.,

the welding community. When the requirements for designing experiments for weld validation tests are well understood by the welding community, then indeed, the computational model should be run without access to the experimental data.

4.3.1 Validation of Transient Thermal Analysis

Because the transient temperature field drives microstructure evolution and thermal expansion, it is critical that the transient temperature field be computed accurately. This is the first stage of the computational validation process. To do this, the research group typically follows the steps below:

1. Adjust the welding start time by matching the time of the first pass peak temperature at the first thermocouple;
2. Adjust the power by matching the peak temperature of the first pass at the first thermocouple;
3. Adjust the welding speed by matching the time of the first pass at the peak temperature at the next thermocouple;
4. Re-adjust the power by matching the peak temperatures of the first pass in the first two thermocouples;
5. Adjust the start dwell time by matching the thermocouple peak temperatures of the first pass;
6. Adjust the convection coefficient function by matching cooling curves after the first pass in all thermo-couples (TCs);
7. Adjust delay times after each pass by matching the times of all peak temperatures at all thermocouples;

8. In some cases, adjust the thermal conductivity.
9. Re-adjust the convection coefficient function by matching cooling curves after all passes finished in all thermocouples.

Most of these steps are needed because accurate experimental data for weld start/end time, start/end position, welding speeds, delay times between weld passes, and so on are not reported. Possible improvements that might be required if the difference between measured and computed temperatures is not acceptable are given below:

1. Use different weld procedures for each pass (power, weld speed) and repeat the previous steps.
2. Use different start dwell times for each pass.
3. Use decreased power at the end of each pass.
4. Change coefficients for each term in the convection coefficient function (we only played with the global scale).
5. Change convection coefficient functions after each pass.

4.3.2 Validation of Microstructure Evolution

The final microstructure can be measured on sections of a weld using metallographic techniques and compared with the computed microstructure. It can be useful to compare measured and computed hardness maps.

4.3.3 Validation of Transient Strain and Displacement Analyses

Once satisfied that the transient temperatures and microstructure analyses are acceptable, the focus turns to the transient displacements and strains.

One of the most important material properties for this stage is the temperature dependent specific volume or density for each phase in the microstructure. In low-alloy steels, ferrite, pearlite, austenite or gamma, bainite, and martensite phases often play important roles. Since the microstructure in low alloy steels can be quite different on heating and cooling, the specific volume at each Gauss point is a function of both temperature and microstructure. This function is history dependent.

The temperature dependent yield stress and hardening modulus are important. Here again, the evolving microstructure makes this non-trivial. The temperature dependent yield stress and hardening modulus of each phase is separately specified followed by computation of the temperature dependent yield stress and hardening modulus at a Gauss point using a rule of mixtures as a function of the fraction of each phase present and the temperature.

4.3.4 Sensitivity of Computed Results to Model Parameters.

The finite element mesh and time step size must be fine enough to be able to resolve transient temperature, microstructure, and displacement fields with sufficient accuracy. Near the weld pool, a mesh sizes in the range of 1 to 10 [mm] and a time step size near 1 second are utilized.

to adjust the input data to improve the agreement between measured and computed results. These steps are explained in the following sections for thermocouples, strain gauges and dial gauge separately.

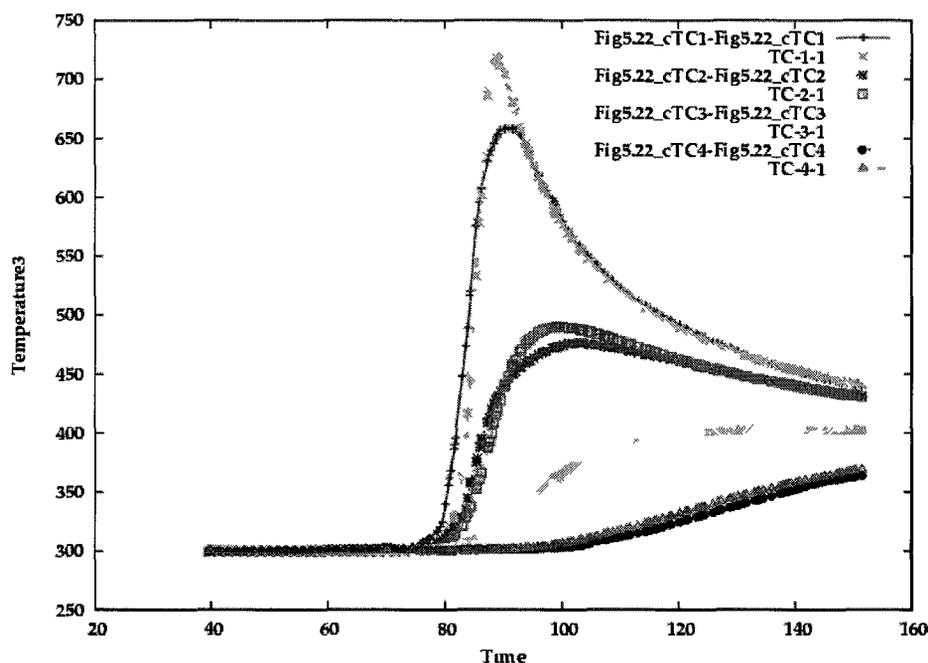


Figure 4.3: Comparison between computed thermal profiles and measured profiles for the four thermocouples. The red and green curves are experimental and computed temperatures for thermocouple 1, the blue and magenta curves are experimental and computed temperatures for thermocouple 2, the blue-green and yellow curves are experimental and computed temperatures for thermocouple 3, the black and orange curves are experimental and computed temperatures for thermocouple 4.

4.4.1 Validation of Transient Temperature Computation

There was no indication of arc weld efficiency given in [27] and therefore different values of arc efficiency were applied to match the peak temperatures measured by each of thermocouples. Figure 4.3 shows the result using an arc efficiency value of 60 %. Values less than 60 % shift the thermal profile below the experimental profile and the values greater than 60 % shift the thermal profile above the experimental profile.

The convection coefficient function was changed by scale factors of 0.5, 0.8, 1.2 and 1.5 and the changes in the transient temperature were not significant.

4.4.2 Validation of Transient Strain Gauge Computation

Figure 4.4 compares the computed results for strain gauges positioned as described in [27] with the experimental data taken from [27]. Thermal strain is fully subtracted from the total strain. Red, dark blue, light blue and black are experimental data for strain gauge 4, 3, 2, and 1 respectively. Green, magenta, yellow and light brown are computed results for strain gauge 4, 3, 2, and 1 respectively. There is a good agreement for strain gauges 1, 2, and 3 but not 4.

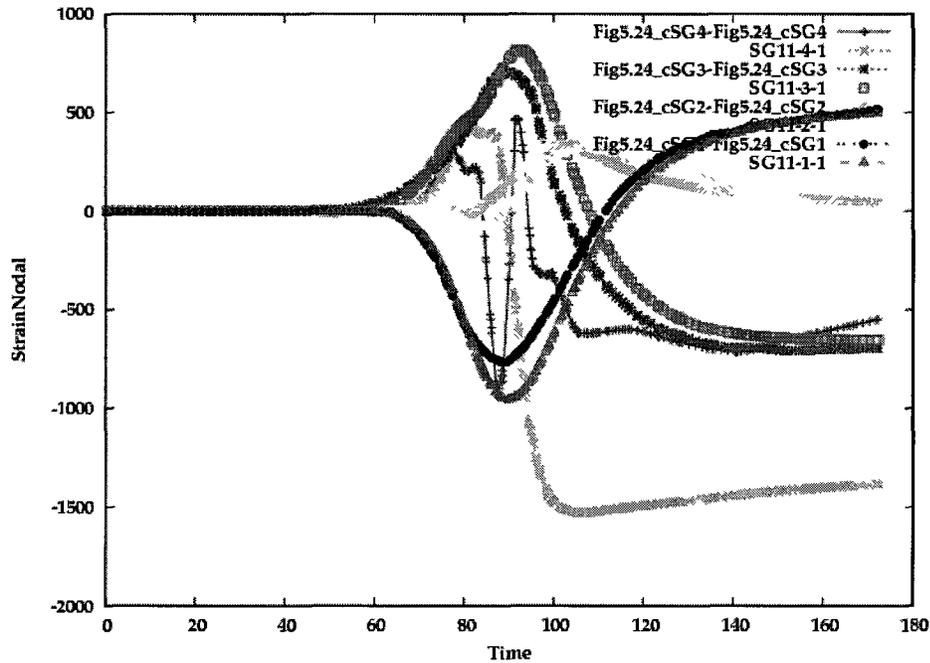


Figure 4.4: Comparison between the computed result for strain gauges with the experimental data taken from [27]. Thermal strain is fully subtracted from the total strain. Red, dark blue, light blue and black are experimental data for strain gauge 4, 3, 2, and 1 respectively. Green, magenta, yellow and light broown are computed results for strain gauge 4, 3, 2, and 1 respectively.

The strain measured with strain gauge 4 was determined to be sensitive to the position of the strain gauge wrt to the weld fusion zone. In [27], the position of the strain gauge is given as a point. In fact, the strain gauge has finite size, possibly it is 10 to 15 [mm] long. The sensitivity of strain gauge 4 to its position wrt to the weld is shown in Fig. 4.5. The results for strain gauge 4 positioned with distances from top edge of 4.13, 3.81, 3.49, 3.17, and 2.85 [mm] are shown in green, dark blue, magenta, light blue and yellow respectively. Experimental data is shown in red.

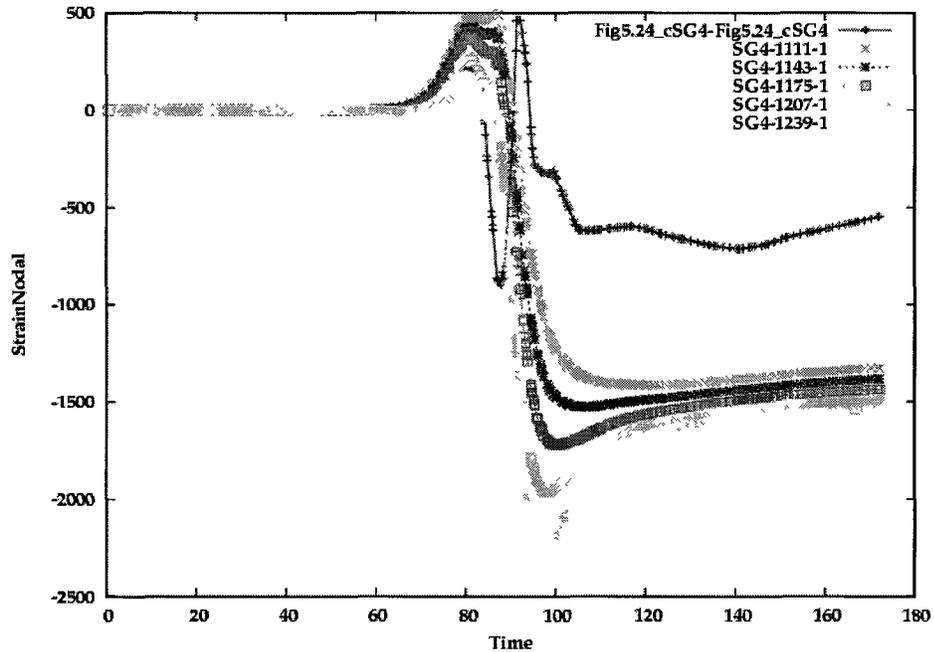


Figure 4.5: Sensitivity of strain gauge 4 to its position distanced from top edge by 4.13, 3.81, 3.49, 3.17, 2.85 [mm] are shown in green, dark blue, magenta, light blue and yellow respectively. Experimental data is shown in red.

As temperatures near the weld pool approach the melting point, the material behavior changes from rate independent thermal-elasto-plastic to rate dependent or visco-thermal-elasto-plastic. When the metal melts the viscosity drops by many orders of magnitude. Because material property data is usually not available for temperatures near the melting point, a cut-off temperature is applied in the stress analysis. This is a low pass filter such that temperatures sent to the stress solver that exceed

this temperature are set to this temperature. To assess the sensitivity of the computed results to the value of the cut-off temperature, values of the cut-off temperatures were varied to temperatures very close but not equal to the melting point.

In addition, the specific volume was perturbed by perturbing the coefficient of thermal expansion. The tests show that increasing the coefficient of thermal expansion (CTE) shifts the end tail of the transient displacement and drops the minimum. Fig. 4.6 compares the settings with cut-off temperature 850 [$^{\circ}K$], convergence criteria 0.000001, maximum number of Newton-Raphson (NR) iterations 10 and increases in the coefficient of thermal expansion by adding 2e-6 from the original temperature dependent values taken from [27]. This is an increase in CTE of about 10 %. The material for filler and body are Al-5052-H32. The temperature dependent properties are taken from [27].

Adjustment for Best Agreement

The best agreement is achieved using a cut off temperature of 850 [$^{\circ}K$] (the melting temperature is 880 [$^{\circ}K$]), convergence criteria 0.000001, max NR 10, and the coefficient of thermal expansion increased by 2e-6 using Al-5052-H32 material for filler metal and body with the properties taken from [27]. The CPU time for this analysis is about 2500 seconds for this setting. Figure 4.7 illustrates this result and the experimental results for comparison.

4.5 Closure of the Analysis

For Masbuchi's test problem, the difference between the measured and computed transient temperatures, displacements and strains were sensitive to the value of the control vector. Experience in performing validation tests for Masbuchi's test problem and for other test problems, suggests that the greatest source of error is introduced

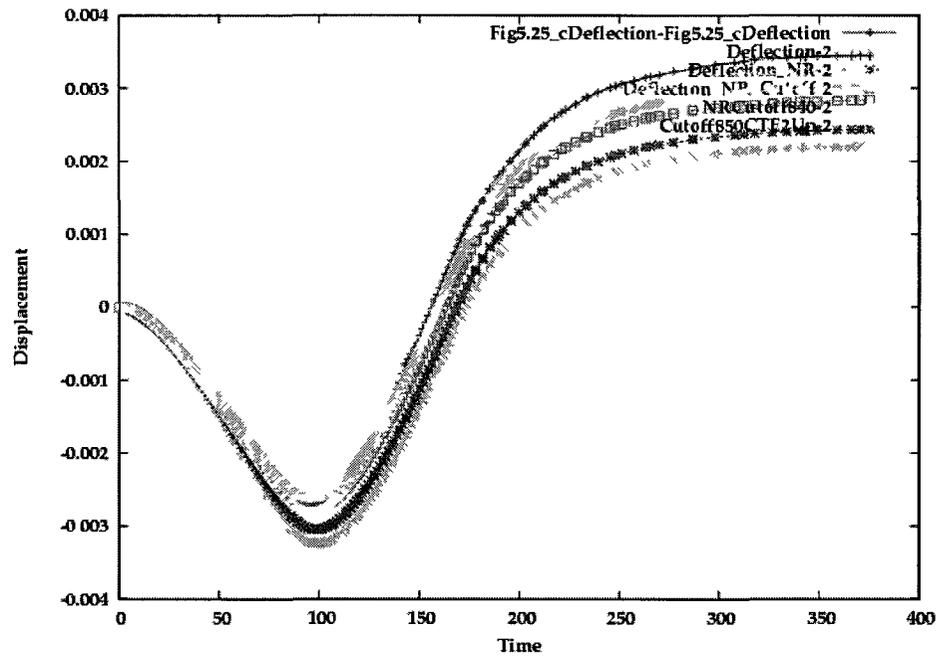


Figure 4.6: Comparison of experimental measured transient displacement, shown in red, with the computed transient displacement. Green shows the setting of cut-off temperature to 700 [$^{\circ}K$], convergence criteria 0.001 and maximum number of NR iterations to 1. Blue shows the setting of cut off temperature to 700 [$^{\circ}K$], convergence criteria 0.000001, maximum number of NR iterations to 10. Magenta shows the setting of cut off temperature to 800 [$^{\circ}K$], convergence criteria 0.000001, maximum number of NR iterations to 10. Light blue shows the setting of cut off temperature to 840 [$^{\circ}K$], convergence criteria 0.000001, maximum number of NR iterations to 10. Yellow shows the setting of cut off temperature to 840 [$^{\circ}K$], convergence criteria 0.000001, maximum number of NR iterations to 10 and an increase in CTE of about 10 %.

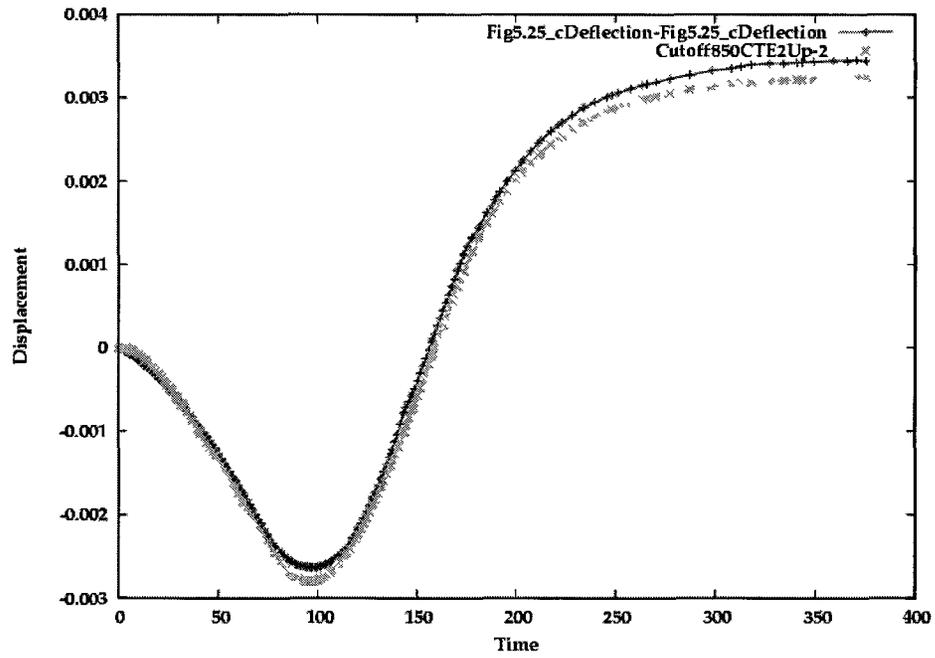


Figure 4.7: Comparison the measured mid-length deflection and computed mid-length deflection with setting of cut-off temperature to 850 [°K] (melting temperature 880 [°K]), convergence criteria to 1e-6, maximum number of NR iterations to 10 and an increased in CTE by adding 2e-6 to the original CTE values taken from [27].

by specifying the control vector, i.e., the parameters that characterize the weld being made such as weld power, weld speed, start times, etc. Errors in geometry, i.e., the difference between the geometry specified in CAD files and the geometry of the real structure being welded can be significant. Once errors in the control vector have been reduced sufficiently, then errors in material composition, temperature dependent material properties and evolution of microstructure are expected to become important. These tend to be more important in the fusion zone and HAZ of the welds.

Chapter 5

A Continuous Design Space Exploration; Space Discretization

Project: Mitigation of Distortion in an Edge-Welded-Bar by Clamping Parameters

The objective is to explore a design space to minimize distortion and evaluate the sensitivity of the distortion of an edge weld on a 152 x 1220 x 12.5 mm bar of Aluminum 5052-H32 wrt clamping parameters. For each point in the design space, a full computational model that includes transient 3D thermal and stress analysis is solved. The bar has no displacement constraints other than rigid body constraints and the resulting camber from running an edge weld, bends the bar. The minimum distortion in this discrete design space is assumed to be the optimal design to minimize the final distortion, i.e., objective function. The design space parameters chosen are clamping parameters, i.e., prescribed displacements, and the release time value in the design space. The bar is then fixed at both ends and subjected to a range of prescribed displacements opposite to the direction of the camber. In the first set of tests the prescribed displacement is applied directly in the middle of the bar and in the second set of tests the displacement field is prescribed as a parabolic

displacement along the full length of the bottom of the bar. In addition to the effect of the prescribed displacement on final distortion is shown to be highly correlated with the delay time at which the prescribed displacement is released after the weld is finished. The best pair of the prescribed value and the release time value in the design space is sought. The distortion and residual stress fields in the mitigated bar with a nodal prescribed displacement in the middle of the bar and the mitigated bar with a parabolic prescribed displacement along the bottom surface of the bar are compared.

5.1 Analysis Background

Distortion in a weld results from inhomogeneous expansion and contraction of the weld metal and adjacent base metal during the heating and cooling cycle of the welding process. Heating and cooling involve significant local volumetric thermal strains that generate permanent plastic deformation. The cause of welding residual stress/distortion is the nonlinear behavior due to this plastic deformation [31] and [32].

Because distortion due to welding can increase fabrication costs and residual stress can reduce the in-service life due to fatigue or corrosion failures, designers would like to minimize their harmful effects. The usual classical approach would be to design a Design-Of-Experiment (DOE) test matrix and perform the physical experiments defined by the test matrix. This is expensive in time and money. Because the DOE test matrix is designed with limited information, understanding or insight of the design space, it may not be the optimal DOE matrix. This paper argues that given such a proposed DOE test matrix, there can be significant benefits from running a multiple computer simulation to evaluate all of the points in the design space that are defined by the DOE test matrix. This capability supports designers to construct and use more efficient DOE matrices.

The experience of the current research group is that the cost of the computer simulation for CWM of industrial welding problems is less than 1% of the cost of the physical experiments and takes less than one week which is much less than the time to perform the physical experiment. An automated machinery for multiple setups and analyses that saves the user's time significantly in computer simulation for CWM of industrial welding problems, can implement larger and advanced DOE matrices in feasible time to support designer-driven optimization and control application of CWM.

In this analysis the design space is chosen to demonstrate this capability. The design parameters are the prebending value and the delay time at which clamps or fixtures are released. Okerblom [9], Vinokurov [10], and Brust [11] discussed different techniques for mitigation of distortion from welding. One of the techniques is to use clamping or prebending. Very good simulation codes are now available to predict the behavior of a welded structure under clamping BCs based on welding heat distribution and stress-strain analysis for every node in domain. However these computer models have largely been limited to cases that analyze a single instance of a weld or welded structure. This paper uses the parametric-design DOE matrix for welds and welded structures that creates a single computer project to solve tens of computer models to explore or map the associated design space specified by a DOE matrix in order to find a minimum distortion in an edge-welded bar constrained by clamping functions. The author emphasizes that the user clicks one 'run' button only once to run all of the projects defined by the DOE matrix.

5.2 Computational Setups

The full computational model that includes thermal and stress analysis is analyzed with VrWeld software [21].

5.2.1 Thermal Analysis

The power density distribution function Q [w/m^3], the ‘Double Ellipsoid’ heat source model [23], was used with the heat source sizes; front, rear, width and depth set to 8, 16, 10, and 8 [mm].

A convection boundary condition generated a boundary flux q [w/m^2] on all external surfaces. This uses ambient temperature of $T_{\text{ambient}} = 300$ [$^{\circ}\text{K}$] and convection coefficient as a function of temperature given in Eq. A.4 for steels

The time step length while welding was chosen so that in one time step the heat source was required to travel one element along the weld path. Filler metal was added as the welding arc moved along the weld path, i.e., the FEM domain changed in each time step during welding. After the weld pass was completed, the time step length was increased exponentially by a factor of 1.2 per time step until the the analysis was halted. The cool down time and the maximum temperature was 3600 seconds and 334 [$^{\circ}\text{K}$] respectively when the analysis halted.

5.2.2 Stress Analysis

The system is solved using a time marching scheme with time step lengths of approximately 3 seconds during welding and an exponentially increasing time step length when welding stops.

5.3 Optimization Problem

When the part is free to deform but rigid body modes are constrained to zero, the final distortion forms a camber in the bar so that the maximum occurs in the middle of the bar as shown in Fig. 5.1. The maximum value of Y -displacement is 3.6 [mm].

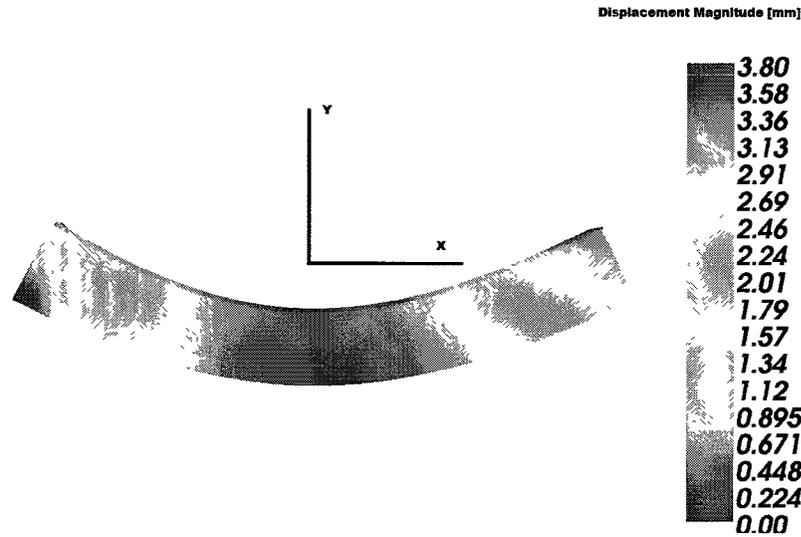


Figure 5.1: Final distortion (50x) when the part is free to deform but rigid body modes are constrained to zero. Red and Yellow lines show the positive X and Y directions.

Prescribing a bending displacement equal and opposite to the camber in the unconstrained bar reduces the final distortion but does not eliminate it entirely. The amount of opposite prescribed camber that minimizes the final camber is one of the design parameters. Not only the magnitude of the prescribed displacement but also the way it is prescribed need to be decided. One strategy is to apply the prescribed displacement in the middle of the bar to a series of nodes along the thickness. Another strategy is to prescribe a parabolic shape of the bar in the direction opposite to the camber to all nodes on the bottom surface of the bar. The effect of prescribed displacement on final distortion is highly correlated with the delay time when the prescribed displacement is released after the weld is finished and it introduces another design parameter in for the clamping technique. Therefore each strategy is solved as an optimization problem and the design parameters are the magnitude of the prescribed displacement and the delay time of the prescribed displacement release after the weld is finished.

5.3.1 Prescribed Displacement in the Middle

In this strategy, a prescribed displacement was applied to a row of nodes in the bottom middle of the bar. Besides rigid body motion fixity, both bottom ends of the bar were fixed with zero Y -displacement. Different values of prescribed displacement were applied to the row of nodes picked at the mid-point of the bar. These prescribed values were held during the welding plus a set of different delay time during the cool down to the ambient temperature. When the prescribed displacement was released, at least one time step computed the equilibrium stress state after releasing the fixities except for rigid body constraints.

The scalar objective function for optimization is taken as the L_∞ norm, i.e., the maximum Y -displacement at the bottom of the bar. The design parameters; values of nodal prescribed displacement and delay time, have a quite large range of possible variation and therefore the design space is discretized by picking 5 values of nodal prescribed displacement and 9 delay times resulting in 45 nodes in the discrete design space. A full factorial DOE matrix including the 45 nodes, was used to give a fully-covered map of the design parameters to the response surface. Table 5.1 shows the DOE matrix employed and the evaluated objective function for each point of design space tested. The result is shown in Fig. 5.2 where each curve illustrates a fitted curve for one prescribed displacement for the whole range of delay time from zero to one hour, i.e, total process time. The fitted curve equations are given in Eq. 5.1, 5.2, 5.3, and 5.4.

Total analysis for each design point takes 48 minutes. Using 4 cores, the DOE matrix finished in 9.6 hours, and the minimum displacement of about zero was computed with no need to refine the discretization of the design space.

Table 5.1: DOE matrix employed for prescribed displacement in the middle of the bar.

Test No	Nodal P	Displ (mm)	Delay Time (mm)	Obj Func (mm)
1		3 80	0	2 04
2		3 80	1	1 80
3		3 80	2	1 71
4		3 80	5	1 65
5		3 80	10	1 60
6		3 80	15	1 57
7		3 80	25	1 51
8		3 80	45	1 46
9		3 80	60	1 46
10		4 75	0	1 59
11		4 75	1	1 33
12		4 75	2	1 22
13		4 75	5	1 16
14		4 75	10	1 11
15		4 75	15	1 07
16		4 75	25	1 01
17		4 75	45	0 96
18		4 75	60	0 96
19		5 70	0	1 16
20		5 70	1	1 00
21		5 70	2	0 77
22		5 70	5	0 68
23		5 70	10	0 62
24		5 70	15	0 58
25		5 70	25	0 51
26		5 70	45	0 46
27		5 70	60	0 46
28		6 67	0	0 71
29		6 67	1	0 44
30		6 67	2	0 27
31		6 67	5	0 16
32		6 67	10	0 10
33		6 67	15	0 06
34		6 67	25	-0 01
35		6 67	45	-0 01
36		6 67	60	-0 01
37		7 60	0	0 20
38		7 60	1	-0 0
39		7 60	2	-0 02
40		7 60	5	-0 04
41		7 60	10	-0 04
42		7 60	15	-0 05
43		7 60	25	-0 05
44		7 60	45	-0 06
45		7 60	60	-0 06

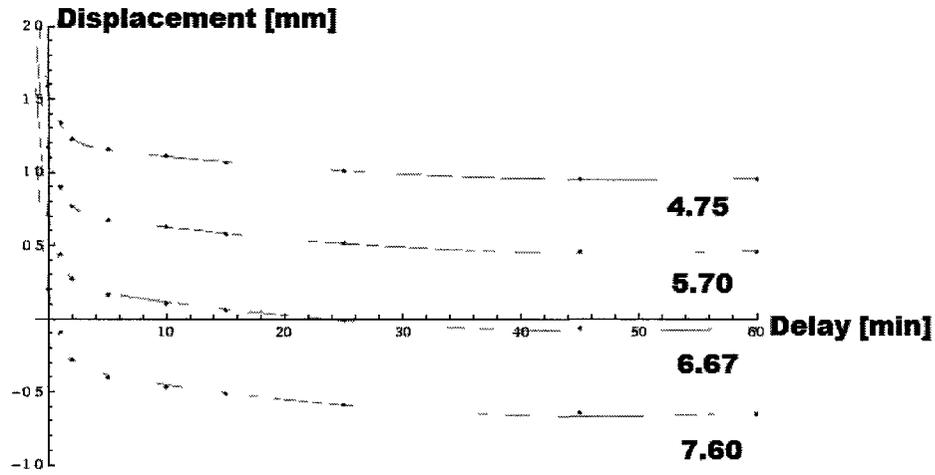


Figure 5.2: Objective function, i.e. maximum final Y -displacement at the end of the process. Each curve illustrates a fitted curve for one prescribed displacement for the range of delay time. Data is given in Table 5.1 and the fitted curve equations are given in Eq. 5.1, 5.2, 5.3, and 5.4 respectively.

$$Y = 0.0001 t^2 - 0.010 t + 0.39 e^{-t} + 1.20 \quad (5.1)$$

$$Y = 0.0001 t^2 - 0.012 t + 0.44 e^{-t} + 0.73 \quad (5.2)$$

$$Y = 0.0001 t^2 - 0.013 t + 0.49 e^{-t} + 0.24 \quad (5.3)$$

$$Y = 0.0001 t^2 - 0.015 t + 0.53 e^{-t} - 0.32 \quad (5.4)$$

where

Y : Objective function in [mm].

t : Delay time in minute.

The curves in Fig. 5.2 were plotted in 3D space and intersected with a flat plane showing zero final displacement as illustrated in Fig. 5.3. The intersection is a curve shown in Fig. 5.4 and the fitted equation to the node picked on the intersection curve, is given in Eq. 5.5. This curve/equation shows the prescribed displacement and the delay time required to achieve zero final displacement.

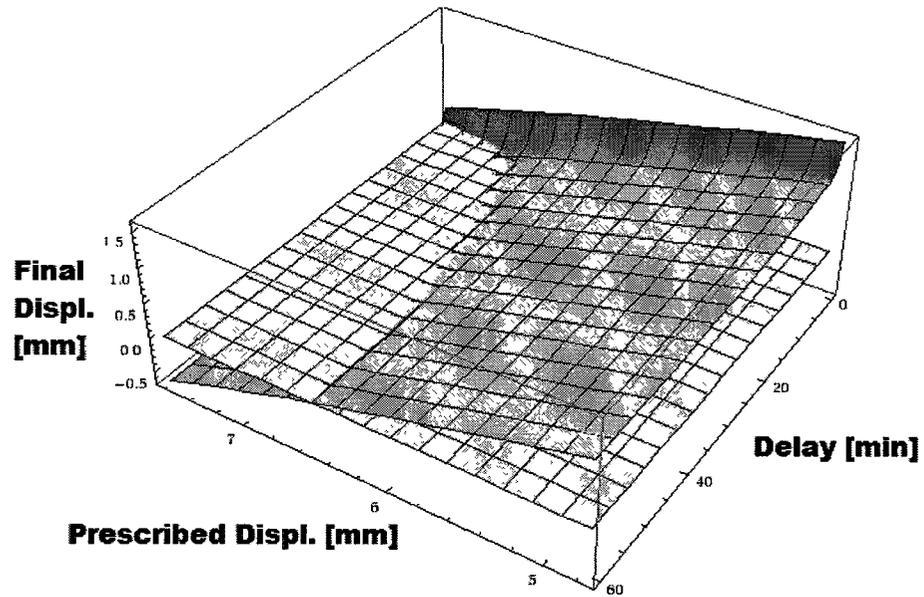


Figure 5.3: Final Y-displacement plane vs the prescribed displacement and the delay time achieved for the nodal prescribed displacement at the end of process and the zero final displacement plane intersects it.

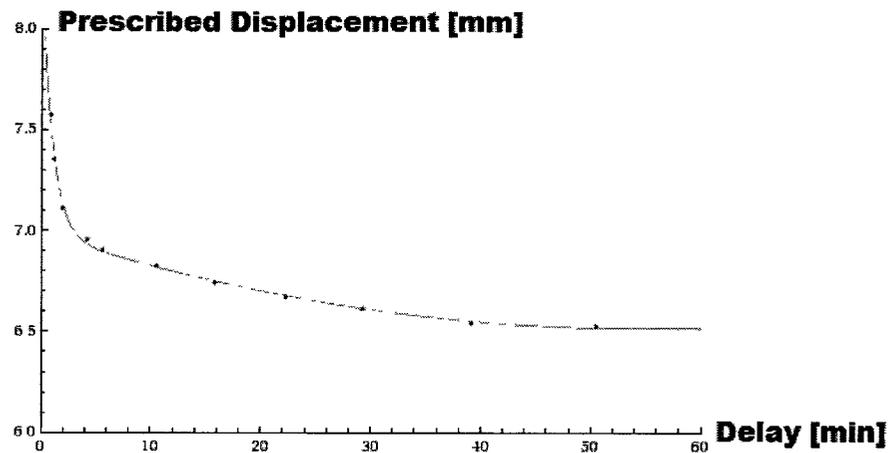


Figure 5.4: The curve shows the prescribed displacement and the delay time required to achieve zero final displacement. The equation is given in 5.5.

$$d = 0.0001 t^2 - 0.017 t + 1.36 e^{-t} + 6.98 \quad (5.5)$$

where

d : Prescribed nodal displacement required to achieve zero final displacement in [mm].

t : Delay time required to achieve zero final displacement in minute.

5.3.2 Parabolic Displacement at the Bottom

In this strategy, a parabolic Y -displacement that is similar but opposite of the Y -displacement in the unconstrained bar, is applied to all nodes at the bottom of the bar. The parabolic function given in Eq. 5.6 is zero at both ends and equals h at the middle of the bar. Different values of h were applied and held during the welding plus a set of different delay times during the cool down to the ambient temperature. When the prescribed displacement was released, at least one time step computed the equilibrium stress state after releasing the fixities except for rigid body constraints.

$$D = h \left(1 + \frac{X}{L} \right) \left(1 - \frac{X}{L} \right) \quad (5.6)$$

where

D : Prescribed displacement function.

h : Maximum displacement in the middle of the bar.

X : Position in the bar varying from $-L$ to L .

L : Bar's half-length.

The DOE matrix is similar to the nodal prescribed displacement test using a meshed space of design parameters with 5 parabolic h values and 9 delay times. Employing 4 cores and having a DOE matrix with the 44 nodes (divisible by 4) saves CPU time. Therefore from the behavior observed in the nodal prescribed displacement test, one node with a low probability of minimum is deleted and a DOE matrix of size 44 is used. This DOE matrix given in Table 5.2, finished in 8.8 hours and the zero minimum was found accurately with no need to refine the mesh of the design parameter space.

Fig. 5.5 illustrates fitted curves for each h value for the whole range of delay times. The fitted curve equations are given in Eq. 5.7, 5.8, 5.9, and 5.10.

$$Y = 0.0001 t^2 - 0.012 t + 0.36 e^{-t} + 1.06 \quad (5.7)$$

$$Y = 0.0001 t^2 - 0.016 t + 0.40 e^{-t} + 0.40 \quad (5.8)$$

$$Y = 0.0001 t^2 - 0.015 t + 0.45 e^{-t} - 0.02 \quad (5.9)$$

$$Y = 0.0001 t^2 - 0.016 t + 0.47 e^{-t} - 0.61 \quad (5.10)$$

where

Y : Maximum final displacement achieved at the end of process in [mm].

t : Delay time in minute.

The curves in Fig. 5.5 were plotted in 3D space and intersected the plane with a flat plane showing zero final displacement as illustrated in Fig. 5.6. The intersection is the line shown in Fig. 5.7 and the fitted equation to the node picked on the intersection line, is given in Eq. 5.11. This line/equation shows h as a function of the the delay time required to achieve zero final displacement.

Table 5.2: DOE matrix employed for parabolic prescribed displacement.

Test No	Parab P	Displ (mm)	Delay Time (min)	Obj Func (mm)
Del		5 70	0	-
1		5 70	1	1 69
2		5 70	2	1 67
3		5 70	5	1 59
4		5 70	10	1 53
5		5 70	15	1 45
6		5 70	25	1 33
7		5 70	45	1 26
8		5 70	60	1 27
9		6 67	0	1 43
10		6 67	1	1 15
11		6 67	2	1 08
12		6 67	5	1 02
13		6 67	10	0 95
14		6 67	15	0 90
15		6 67	25	0 81
16		6 67	45	0 75
17		6 67	60	0 75
18		7 86	0	0 82
19		7 86	1	0 51
20		7 86	2	0 41
21		7 86	5	0 34
22		7 86	10	0 26
23		7 86	15	0 20
24		7 86	25	0 12
25		7 86	45	-0 01
26		7 86	60	0 01
27		8 55	0	0 43
28		8 55	1	0 12
29		8 55	2	0 01
30		8 55	5	-0 01
31		8 55	10	-0 14
32		8 55	15	-0 21
33		8 55	25	-0 30
34		8 55	45	-0 37
35		8 55	60	-0 37
36		9 50	0	-0 13
37		9 50	1	-0 46
38		9 50	2	-0 57
39		9 50	5	-0 67
40		9 50	10	-0 75
41		9 50	15	-0 81
42		9 50	25	-0 90
43		9 50	45	-0 98
44		9 50	60	-0 98

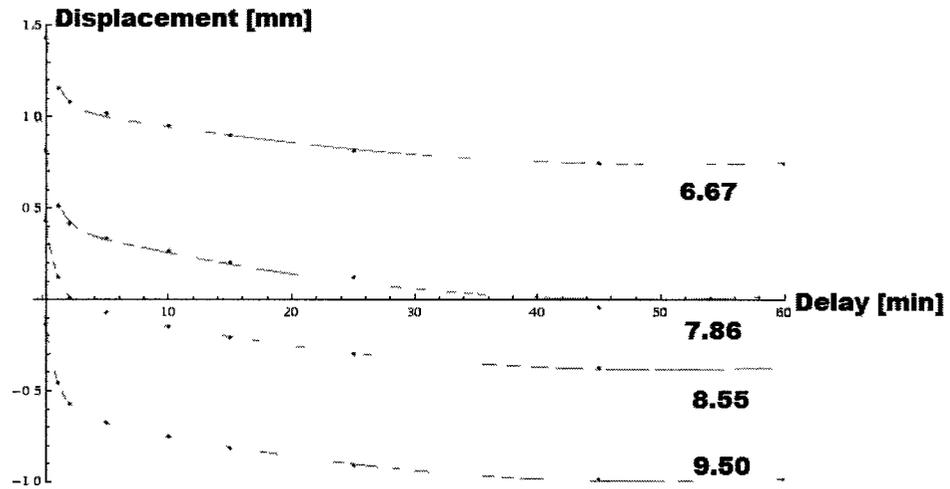


Figure 5.5: Objective function, i.e. maximum final Y -displacement at the end of the process. Each curve illustrates a fitted curve for each h values. Data is given in Table 5.2 and the fitted curve equations are given in Eq. 5.7, Eq. 5.8, Eq. 5.9, and Eq. 5.10 respectively.

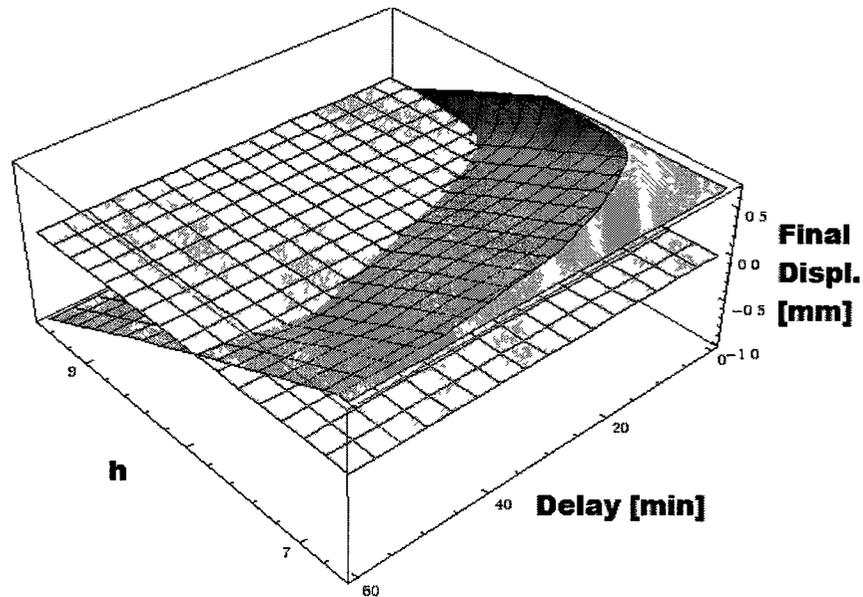


Figure 5.6: Final Y -displacement plane vs h and the delay time achieved for the parabolic prescribed displacement at the end of process and the zero final displacement plane intersects it.

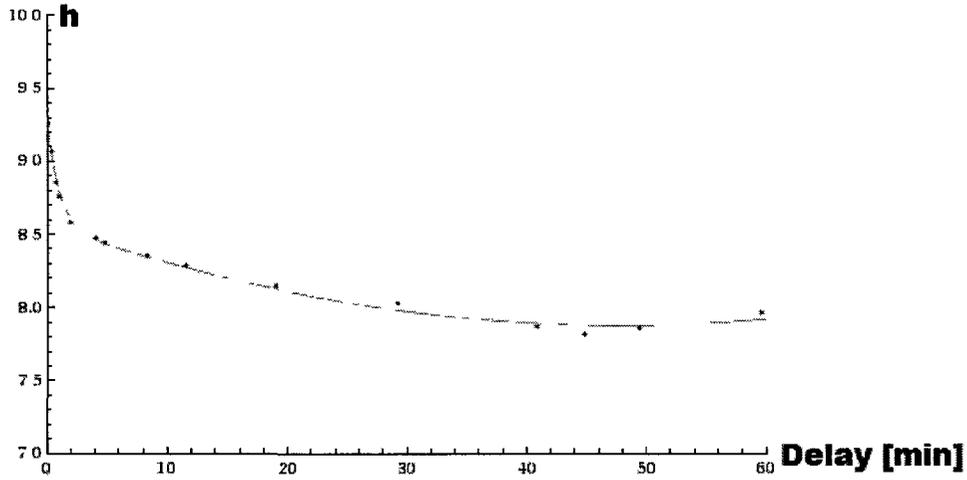


Figure 5.7: Curve shows h (mm) and the delay time required to achieve zero final displacement. The equation is given in 5.11.

$$h = 0.0003 t^2 - 0.029 t + 0.73 e^{-t} - 8.56 \quad (5.11)$$

where

h : Prescribed h required to achieve zero final displacement in [mm].

t : Delay time required to achieve zero final displacement in minute.

5.4 CWM Results and Comparison

Fig. 5.8 compares the final deflection for the two mitigation strategies discussed. Distance is from the left bottom corner to the right bottom corner of the bar and the units are meters. Deflection is minimized with the parabolic prebending. For prebending with prescribed deflections applied at the mid-point of the lower edge and the end points of the lower edges, the reaction forces or Lagrange multipliers are concentrated loads that generate local plastic deformation. These results show that prebending strategy is very effective technique because the maximum final distortion

is reduced from 3.8 [mm], Fig. 5.1, to less than 0.1 [mm]. The parabolic prebending can be a better choice because the bar is closer to a straight bar at the end.

Fig. 5.9 compares the longitudinal residual stress in the bar for the two mitigation strategies after welding is complete. Residual stress is plotted for a line normal to the weld from the top edge to the bottom edge of the bar at the mid-length of the bar. Units are [Pa] and [m] for stress and distance respectively. For prebending with prescribed deflections applied at the mid-point of the lower edge and the end points of the lower edges the reaction forces or Lagrange multipliers generate plastic strain and a very high residual stress at the bottom edge of the bar. For this case, the longitudinal residual stress in the weld has the highest compressive stress of the mitigation methods.

It would be a simple matter to solve a DOE matrix to minimize longitudinal residual stress after welding was complete and the bar cooled down to room temperature if one chooses an objective function. The question is what objective function to choose? Two possible choices are the maximum tensile residual stress or the integral of the square of the longitudinal stress over the cross-section $\int_0^W \sigma_{xx}^2 dW$ where W is the width of the bar. There are many other possible objective functions. Which metric is preferred would depend on the requirements that the designer is trying to satisfy. It is also possible to have multiple objective functions and compute the Pareto optimal solutions.

5.5 Closure of the Analysis

A DOE matrix is used to define a set of points in the design space to be evaluated in order to find a minimum distortion in the bar for various clamping parameters. The set of points in the design space defined by the DOE matrix are evaluated by the software as a single problem with the DOE matrix as an input file to the software.

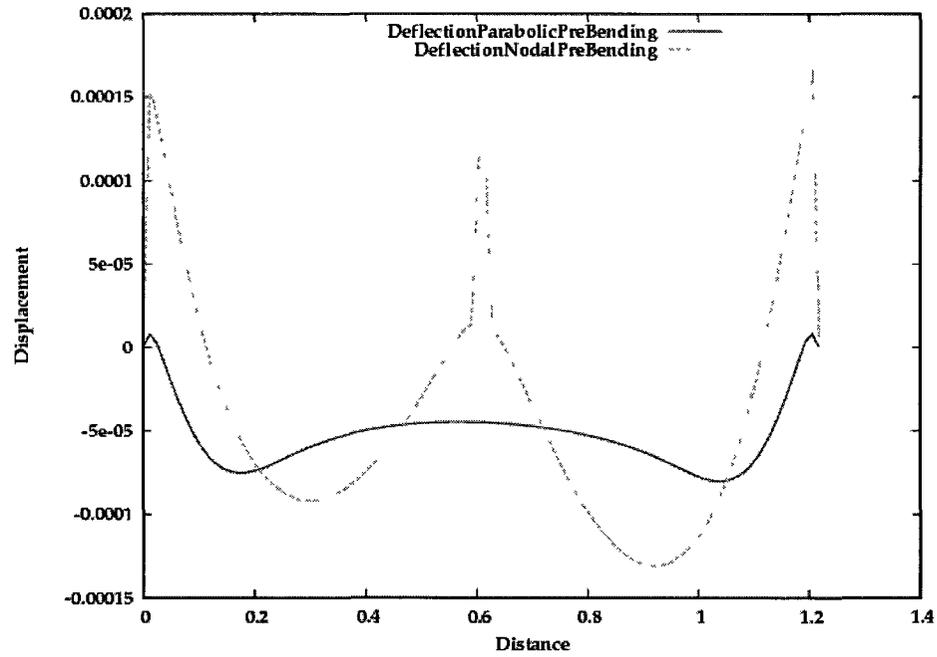


Figure 5.8: Final deflection from the two mitigation strategies employed. Distance is from the left bottom corner to the right bottom corner of the bar. Units are [m].

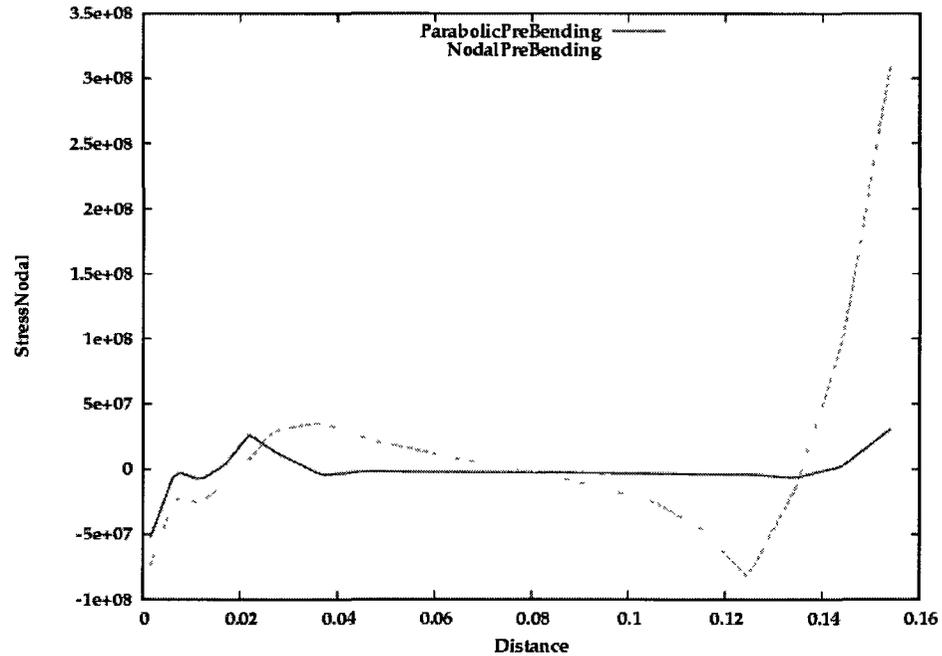


Figure 5.9: Longitudinal residual stress, σ_{xx} , plotted along a line normal to the weld, Y -direction, from the top edge to the bottom edge of the bar in the mid-length. Units are [Pa] and [m].

This is quite different from the user using the DOE matrix to separately create and separately solve one project for each row of the DOE matrix.

The user sets up only one reference or base project and one DOE matrix with 45 design points. This ran as a single project that analyzed all 45 design points in 9.6 CPU hours on a single core of a 3.3 GHZ Intel quad-core processor. With the same base project, a second DOE matrix was created with 44 design points. This analysis required 8.8 CPU hours. The user spends almost no time setting up the analysis for any design point other than the design point for the base or reference project. In addition, the results and post-processing are all structured for each DOE matrix. In the author's experience, it would takes weeks of an expert user's time to set up the 89 projects for these two DOE matrices. Furthermore, managing post-processing would be difficult, time consuming and prone to error.

For each point in the design space, the transient thermal and displacement fields are computed. This analysis discretizes the design space to explore or map the associated design space in order to find a minimum distortion by clamping parameters. This methodology could be applied to any welded structure to explore any set of points in a design space specified by a DOE matrix as a single project in which the variables are parameters of the computational model.

Chapter 6

A Continuous Design Space Exploration; Direct Searching in the Space

Project: A Direct-Search CWM Optimization Using Least-Square Approximation to Mitigate an Edge-Welded-Bar Distortion by Side Heat Source

This analysis focuses on multiple DOE matrices to explore or map the associated design space specified by the DOE matrices to a response surface. Using a framework for exploring a design space in CWM, a recent direct-search algorithm from Kolda, Lewis and Torczon [33] is modified to use a least-square approximation to improve the method of following a path to the minimum in the algorithm. To compare the algorithms, an optimization problem on a 152 x 1220 x 12.5 [mm] bar of Aluminum 5052-H32 is solved to minimize the weld distortion mitigated by a side heating technique. The optimization problem is to find the best point in the space of side heater's design parameters of this technique: power, heated area, longitudinal and transverse distance from weld such that the final distortion is as low as possible (minimized). This optimization problem is constrained to keep the stress level generated by the side heaters, in the elastic region to avoid adding an additional permanent plastic

strain to the bar. The two algorithms are compared for the number of iterations, size of DOE matrices required and CPU time to find the minimum.

6.1 Analysis Background

In general, optimization refers to finding/choosing the best solution from some set of available alternatives. In other words, finding best-available value(s) of some objective functions in a given domain [34]. This is a scalar function f that depends on one or more independent variables and the objective is to find the value of those variables where f is a minimum [17]. There is no perfect algorithm for all forms of nonlinear optimization problems and it is strongly recommended that more than one algorithm be tried and compared [17].

Optimization problems start with one (or more) scalar objective function that determines how good a solution/result is. This scalar-valued function is basically defined by the user to characterize a prefer behavior of a system. The specific form of such function comes from the nature of the problem and creativity of user. However, a p-norm could be a good function for many cases [13]. Simply, an L_p norm is defined as Eq. 6.1 and $p = 1, 2$ and ∞ as Eq. 6.2, 6.3, and 6.4, are frequently used.

$$L_p = \|x\|_p = (|x_1|^p + |x_2|^p + |x_3|^p + \dots + |x_n|^p)^{1/p} \quad (6.1)$$

$$L_1 = \|x\|_1 = (|x_1| + |x_2| + |x_3| + \dots + |x_n|) \quad (6.2)$$

$$L_2 = \|x\|_2 = (|x_1|^2 + |x_2|^2 + |x_3|^2 + \dots + |x_n|^2)^{1/2} \quad (6.3)$$

$$L_\infty = \|x\|_\infty = \text{Max}\{|x_1|, |x_2|, |x_3|, \dots, |x_n|\} \quad (6.4)$$

6.2 Direct-Search Algorithm

Kolda, Lewis, and Torczon have proposed a basic algorithm called “*Compass search*” and an advanced one called “*GSS*” for direct-search optimization in [35]. Their ideas are summarized below.

Given an objective function, a DOE matrix formed for a starting point with a trial step size, and a supporting framework:

1. Run DOE matrix and assess results.
2. If there is a new minimum with a sufficient decrease in the results, then

Move to the minimum as a new trial point and go to 1.

3. If there is no new minimum point in the results, then

Test for convergence.

If no convergence, refine the trial step size and go to 1.

If convergence, write report and stop.

This pseudo-code is employed in our frame work, and an improved method is analyzed to follow more efficiently the path to a minimum using least-square approximation.

6.3 Least-Square Algorithm

Least squares methods, generally, approximate over-determined systems of equations such that the sum of the squares of the residuals becomes minimum. This method is the basis for the regression models that are frequently used to approximate the behavior of a system as a mathematical function of independent variables. The least-square method was employed to approximate the behavior of the objective function

around the trial point and the regression coefficients, $\hat{\beta}$, are calculated by Eq. 6.5 for $\hat{Y} = X\hat{\beta}$ where the $\hat{Y}_{l \times 1}$ and $Y_{l \times 1}$ are the approximation and deterministic vector of n values of the objective function respectively, $X_{l \times k}$ is the matrix of l sample points in a design space and k terms in the regression model, $\hat{\beta}_{k \times 1}$ is computed from Eq. 6.5.

$$\hat{\beta} = (X^T X)^{-1} X^T Y \quad (6.5)$$

The idea of using least-square approximation in direct-search optimization, is presented as the following pseudo-code:

Given an objective function, starting point, parameter step-sizes and parameter bounds construct a mesh or grid to discretize the design space:

1. Run DOE matrix and assess results.
2. Construct the least-square regression model for the objective function on the assessed results.
3. Test for convergence, then

If converged, write report and stop.

If not, find values of parameters for the minimum of the regression equation.

If the minimum from regression approximation is smaller than accurate results in the DOE matrix then run the configuration of the regression-found minimum using the accurate model.

If the accurate model and the regression model agree within a certain tolerance, move to the new configuration.

If not, move to the minimum in the regression-tested region. Find the

closest point on the mesh (or 2R-refined mesh ¹) and perturb based on the gradient computed from the least-square approximation on the node picked.

4. Go to 1

Both algorithms; the original direct-search and the least-square direct-search, are employed in this analysis to solve an optimization problem to mitigate the distortion of an edge welded Aluminum bar using a side-heating technique. The optimization problem is to find the point in the space of side heater design parameters that minimizes the final distortion. The two algorithms are compared with respect to the number of iterations, size of DOE matrices required and CPU time to find the minimum in the discussion section.

6.4 Edge-Welded Bar Setup

The test setup is an edge weld on a 152 x 1220 x 12.5 [mm] bar of Aluminium 5052-H32. The full computational model that includes transient thermal and stress analysis is analyzed by VrWeld software [21]. CWM validation is presented in chapter 4 by comparison with an experimental data measured carefully by Masubuchi in [27].

The mesh employed (Fig. 6.1) has 6600 8-node brick elements and 9680 nodes. The material was Aluminum 5052-H32 alloy with chemical composition Al 96.7, Mg 2.5, Cr 0.25, Cu max 0.1, Fe max 0.4, Mn max 0.1, Si max 0.25, Zn max 0.1 [Wt %]. The temperature-dependent material properties of Al 5052-H32 were given in [27] and this data was employed in the analysis of this test. The gas metal-arc-welding process was employed to weld the specimen and the welding parameters were current 260 [amperes], voltage 23 [volts], travel speed 7.34 [mm/s], filler metal Al-4043 with 1.6 [mm] wire diameter, wire feed speed 170 [mm/s], and the shielding gas was

¹One point is added in the middle of the existing mesh line that connects two points on the current mesh.

Argon. The specimen was allowed to cool to ambient temperature after welding was completed.

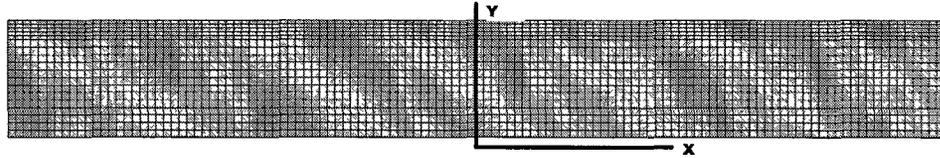


Figure 6.1: A 2D view of the 3D mesh employed in the analysis.

The stress analysis shows a significant Y-deflection illustrated in Fig. 6.2 and the goal is to mitigate this deflection in order to get as straight a bar as possible at the end of the process.

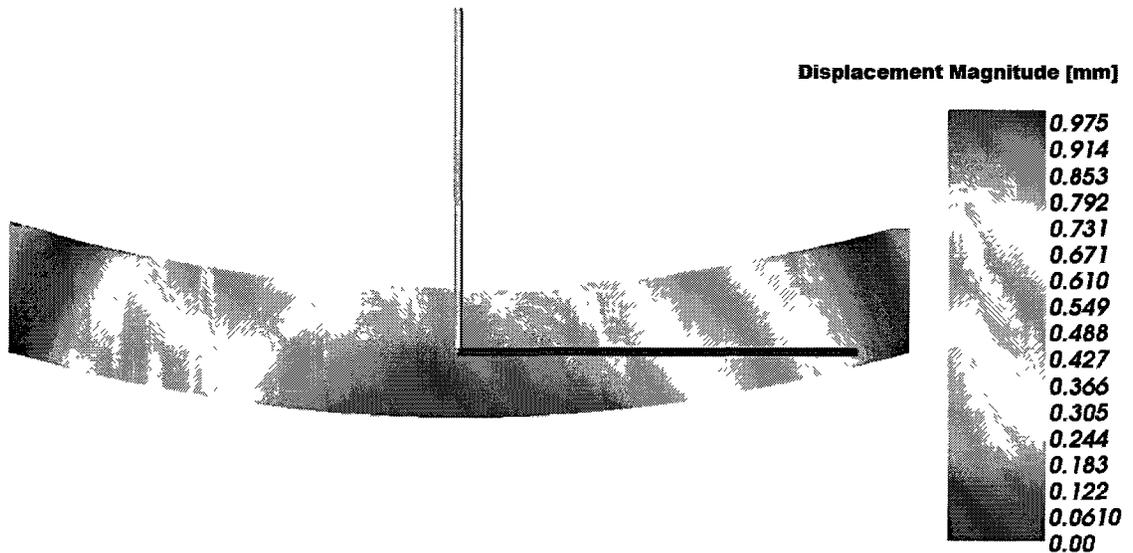


Figure 6.2: Deflection in Y direction at the end of the process (x85). Red is X direction and yellow is Y direction.

Okerblom [9], Vinokurov [10], and Brust [11] discussed different techniques for mitigation of distortion from welding. One of the techniques is to apply a transient thermal tension by side heaters [6]. This technique introduces a significant tension around the weld. The side heater's power, heated area, the distance from the weld either longitudinal or transversal are the design parameters for this technique.

The side heater source is characterized by a double ellipsoid model [23] moving parallel to the weld path. The power is computed from ηVI ; side heater efficiency, current and voltage. Power is varied by changing η , efficiency, from 0.2 to 0.7 using fixed I and V equal to 260 [ampers] and 23 [volts]. Four semi-axes lengths of the parameters of the double ellipsoid geometry are assumed equal and therefore form a sphere. In effect, the area formed by the intersection of this sphere and the surface of the bar is the area that absorbs the power and therefore is one of the side heating parameters. This parameter, area, was characterized by a single value that is the radius of the sphere, R. This parameter ranges from 10 to 70 [mm]. The quasi-transient position of the side heater wrt the weld, can be moved ahead/behind the arc or shifted closer or farther from the weld path. The origin of the coordinate system was placed on the bar's centerline and exactly below the weld tip as shown in Fig. 6.3. The relative position of the side heater therefore can move in the X or Y direction. Finally, the side heater design parameters are; η , R, (X, Y), denote power, area's radius, longitudinal, and transverse shift respectively.

The side heating could add plastic strain to the bar if the power density is high. To avoid forming such plastic strain, power is constrained to be in the gray area in Fig. 6.4. This plot is drawn based on the plastic strain computed by FEM analyses when the side heater is applied with no weld. The gray area shows the no-plastic-strain zone. The analyses show that if the maximum temperature in the side heater stays below 480 [$^{\circ}K$], the plastic strain does not form. Since thermal analysis requires much less CPU time than stress analysis, the upper limit of 480 [$^{\circ}K$] can be used in an optimization algorithm to apply this constraint.

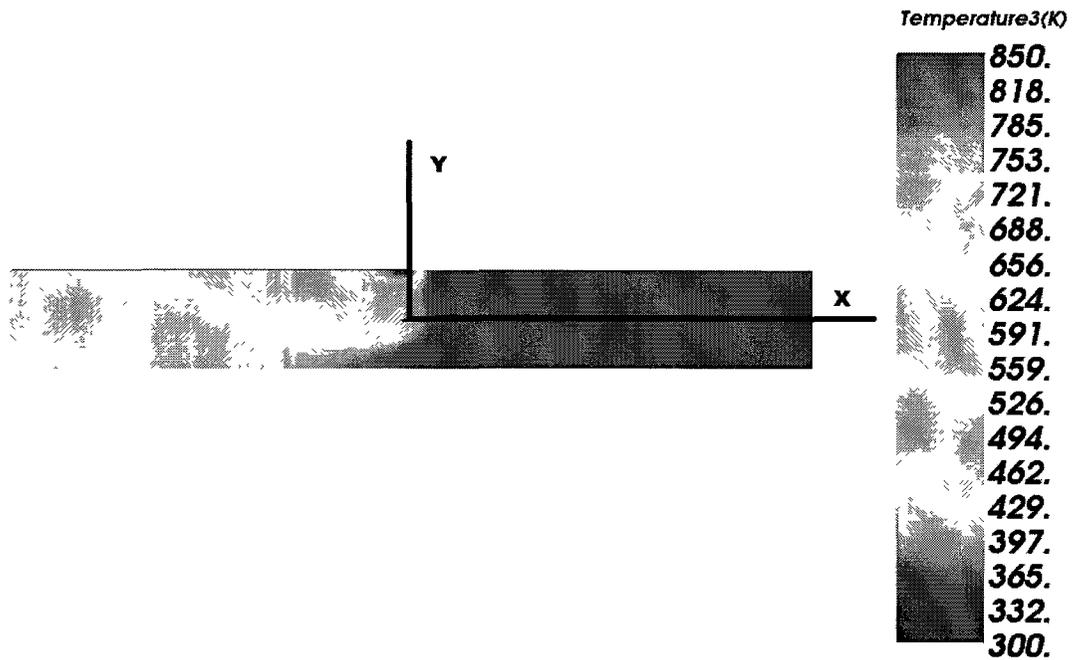


Figure 6.3: Origin of the coordinate system used in the analysis. Red is X direction and yellow is Y direction.

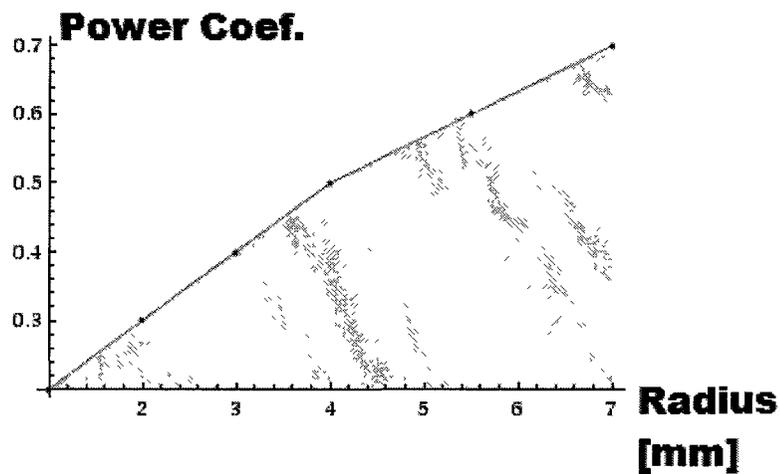


Figure 6.4: Constraint showing the allowable region for two side heater parameters: power and area. Nodes on the gray zone have the maximum temperature in a side heater below 480 [°K] and therefore generate no plastic strain.

6.5 Computational Setups

The full computational model that includes thermal and stress analysis is analyzed with VrWeld software [21].

6.5.1 Thermal Analysis

The power density distribution function Q [w/m^3], the ‘Double Ellipsoid’ heat source model [23], was used with the heat source sizes; front, rear, width and depth set to 8, 16, 10 and 8 [mm].

A convection boundary condition generated a boundary flux q [w/m^2] on all external surfaces. This uses ambient temperature of $T_{\text{ambient}} = 300$ [$^{\circ}\text{K}$] and convection coefficient as a function of temperature given in Eq. A.5.

The time step length while welding was chosen so that in one time step the heat source was required to travel one element along the weld path. Filler metal was added as the welding arc moved along the weld path, i.e., the FEM domain changed in each time step during welding. After the weld pass was completed, the time step length was increased exponentially by a factor of 1.2 per time step until the the analysis was halted. The cool down time and the maximum temperature was 3600 seconds and 334 [$^{\circ}\text{K}$] respectively when the analysis was halted.

6.5.2 Stress Analysis

The displacement boundary conditions for the test removed the rigid body modes by constraining the bottom left edge to zero displacement and also by constraining the bottom right edge to zero transverse motion and zero vertical motion but allowing horizontal translation.

The system is solved using a time marching scheme with time step lengths of approximately 3 seconds during welding and an exponentially increasing time step

length when welding stops.

6.6 Optimization Problem for the Edge-Welded Bar

As an optimization problem for the edge-welded bar, the objective is to obtain the values of η , R , X , and Y such that the final distortion is minimized. Since the main deflection is in the normal direction to the weld (Y direction in our setup), the scalar objective function is picked as the L_1 norm of the final Y displacement along the bottom of the bar at the end of the process. If the position of nodes is set on undeformed geometry (original part before welding) to zero, the new position on deformed geometry (distorted part after welding) then can be defined as x in Eq. 6.2 showing the nodal deformation. The total number of nodes in the L_1 norm calculation was 100 along the bottom of the bar. The next step in the optimization is to define an initial DOE matrix.

Taguchi [19] introduced several DOE matrices to explore the space of design parameters with a minimal number of tests and there are a great number of optimization efforts using his philosophy [36]. Although Taguchi's approach is an efficient method, a few concerns have been raised. Some of these concerns relate to the absence of higher-order interactions of design parameters. For these reasons, other approaches to carry out robust parameter design have been suggested including response modeling [37]. Since optimization algorithms require a start point and having a good start point leads to a better optimization, this suggested the optimization be started by a Taguchi's DOE matrix to screen our parameters wrt the sensitivity of the objective function to each parameter. Therefore, an L_9 Taguchi DOE matrix was employed as given in Table 6.1 and the range of variations conforms with the constraint shown in

Fig. 6.4.

Table 6.1: L_9 Taguchi DOE matrix employed to screen the optimization parameters wrt the sensitivity of the objective function to each parameters.

No.	R [m]	η	Y [m]	X [m]	Obj. Func.
1	0.04	0.3	-0.0162	-0.0508	0.099
2	0.04	0.4	0.0	0.0	0.090
3	0.04	0.5	0.0162	0.0508	0.148
4	0.05	0.3	0.0	50.8	0.134
5	0.05	0.4	0.0162	-0.0508	0.080
6	0.05	0.5	-0.0162	0.0	0.102
7	0.06	0.3	0.0162	0.0	0.091
8	0.06	0.4	-0.0162	0.0508	0.142
9	0.06	0.5	0.0	-0.0508	0.056

The L_9 results show that the sensitivity of the objective function to the radius is high. This suggests minimizing the objective function, first, wrt to the radius separately. However the radius is linked to the power by the constraint (Fig. 6.4). Therefore it is continued to minimize the objective function, wrt to the radius and power respecting the constraint. A set of 4 points (0.04 m, 0.5),(0.04 m, 0.3),(0.02 m, 0.3),(0.03 m, 0.4) was picked on the allowable area of the constraint and Y and X were perturbed with 3 levels around these points. This gives a second DOE matrix with 36 tests as listed in Table 6.2. The reason for picking this set of points was that the tests numbered 10 and 27 were available from the previous stage and CPU time was saved by reading the result. However some other sets of points were tested and led to the same conclusion.

Table 6.2: DOE matrix used to minimize the objective function, wrt to the radius and power respecting the constraint.

No.	R [m]	η	Y [m]	X [m]	Obj. Func.
1	0.03	0.4	0.0162	0.0508	0.112
2			0.0162	0.0	0.069
3			0.0162	-0.0508	0.071
4			0.0	0.0508	0.117
5			0.0	0.0	0.086
6			0.0	-0.0508	0.080
7			-0.0162	0.0508	0.145
8			-0.0162	0.0	0.127
9			-0.0162	-0.0508	0.129
10	0.04	0.5	0.0162	0.0508	0.099
11			0.0162	0.0	0.050
12			0.0162	-0.0508	0.057
13			0.0	0.0508	0.103
14			0.0	0.0	0.070
15			0.0	-0.0508	0.058
16			-0.0162	0.0508	0.135
17			-0.0162	0.0	0.114
18			-0.0162	-0.0508	0.115
19	0.04	0.3	0.0162	0.0508	0.137
20			0.0162	0.0	0.107
21			0.0162	-0.0508	0.111
22			0.0	0.0508	0.140
23			0.0	0.0	0.116
24			0.0	-0.0508	0.114
25			-0.0162	0.0508	0.160
26			-0.0162	0.0	0.146
27			-0.0162	-0.0508	0.148
28	0.02	0.3	0.0162	0.0508	0.129
29			0.0162	0.0	0.094
30			0.0162	-0.0508	0.093
31			0.0	508	0.134
32			0.0	0.0	0.108
33			0.0	-0.0508	0.106
34			-0.0162	0.0508	0.157
35			-0.0162	0.0	0.142
36			-0.0162	-0.0508	0.145

The objective function values in Table 6.2 show that there is no significant interactions between parameters (X, Y) and parameters (R, η) because perturbation in (X, Y) gives the similar behavior for 4 settings of R and η in terms of sorting the objective function ascending/descending. In other words, (X, Y) perturbation has a similar response surfaces shifted in different R, η . This DOE matrix indicates the optimum configuration of R and η is on a boundary line of the constraint toward the higher value of R and η . The upper bound of the radius is reduced to 6 [cm] to be narrower than the bar (7 [cm]) and therefore the optimum for the R and η is R = 6 [cm] and $\eta = 0.6$, i.e., 3588 J.

In order to find the optimum value of Y and X, both algorithms are implemented:
a) the original direct-search algorithm, b) the least-square direct-search algorithm.

6.6.1 Original Direct-Search Algorithm

The discussed original direct-search pseudo-code implemented for the starting point of Y and X equal zero, the trial step size of 0.0162 [m] and 0.0508 [m] for Y and X and the minimum step size of 0.002 [m] and 0.003 [m] for Y and X. The implementation generates the sequence of data presented in Table 6.3. This is illustrated graphically in Fig. 6.5 to show the path followed by this method to the minimum.

6.6.2 Least-Square Direct-Search Algorithm

The discussed least-square direct-search pseudo-code implemented for the starting point of Y and X equal zero, the trial step size of ± 0.02 [m] Y and X and the minimum step size of 0.002 [m] and 0.003 [m] for Y and X. The implementation of least-square direct-search algorithm results the sequence of data presented in Table 6.4. The least-square regression Eq.s 6.6, 6.7 and 6.8 are constructed and used for the approximation. This is illustrated graphically in Fig. 6.6 to show the path followed by

Table 6.3: Data table for the original direct-search algorithm tested and the decision at the end of each iteration. C indicates the result copied from previous runs to avoid repeating the analysis.

Iteration	Run No	Y [m]	X [m]	Obj Func	Decision	
1	1	0 0162	0 0508	0 1070		
	2	0 0162	0 0	0 0625		
	3	0 0162	-0 0508	0 0642		
	4	0 0	0 0508	0 1012		
	5	0 0	0 0	0 0681		
	6	0 0	-0 0508	0.0606		
	7	-0 0162	0 0508	0 1299		
	8	-0 0162	0 0	0 1100		
	9	-0 0162	-0 0508	0 1113		Move
2	C	0 0162	0 0	0 0625		
	C	0 0162	-0 0508	0 0642		
	10	0 0162	-0 1016	0 0960		
	C	0 0	0 0	0 0681		
	C	0 0	-0 0508	0.0606		
	11	0 0	-0 1016	0 0703		
	C	-0 0162	0 0	0 1100		
	C	-0 0162	-0 0508	0 1113		
13	-0 0162	-0 1016	0 1183	Refine		
3	14	0 0081	-0 0254	0.0512		
	15	0 0081	-0 0508	0 0518		
	16	0 0081	-0 0762	0 0618		
	17	0 0	-0 0254	0 0623		
	C	0 0	-0 0508	0 0606		
	19	0 0	-0 0762	0 0602		
	20	-0 0081	-0 0254	0 0623		
	21	-0 0081	-0 0508	0 0820		
	22	-0 0081	-0 0762	0 0826		Move
	4	23	0 0122	-0 0254		0.0495
24		0 0081	-0 0145	0 0526		
C		0 0081	-0 0254	0 0512		
26		0 0081	-0 0363	0 0514		
27		0 0041	-0 0254	0 0554	Move	
5	28	0 0162	-0 0254	0 0524		
	29	0 0122	-0 0145	0 0510		
	C	0 0122	-0 0254	0.0495		
	30	0 0122	-0 0363	0 0509		
	31	0 0081	-0 0254	0 0512		Refine
6	32	0 0145	-0 0254	0 0504		
	33	0 0122	-0 0200	0 0499		
	C	0 0122	-0 0254	0.0495		
	34	0 0122	-0 0309	0 0501		
	35	0 0101	-0 0254	0 0524		Stop

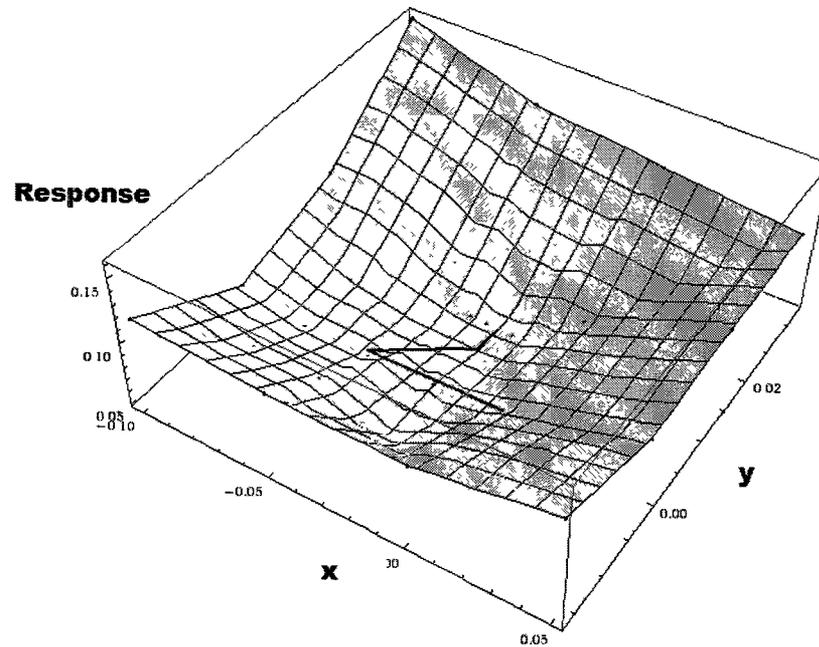


Figure 6.5: The original direct-search algorithm results (Table 6.3) is illustrated graphically to show the path followed by this method to the minimum.

the least-square code to the minimum and compared with the original direct-search code.

$$Reg_1 = 8.65 Y^2 + 0.43 Y + 73.50 X^2 - 1.31 X + 0.07 \quad (6.6)$$

$$Reg_2 = 11.45 Y^2 + 0.58 Y + 107.77 X^2 - 2.31 X + 0.07 \quad (6.7)$$

$$Reg_3 = 14.59 Y^2 + 0.74 Y + 92.98 X^2 - 2.16 X + 0.07 \quad (6.8)$$

6.7 Comparison and Discussion

Both algorithms find the same minimum, i.e., $Y = -0.025$ and $X = 0.012$ [mm]. The difference is the number of iterations required to find the minimum. The original

Table 6.4: Data table for the least-square direct-search algorithm tested and the decision at the end of each iteration. C indicates the result copied from previous runs to avoid repeating the analysis.

Iteration	Run No.	Y [m]	X [m]	Obj. Func.	Decision
1	1	0.02	0.0	0.0714	Reg. Eq. 6.6
	2	0.0	0.02	0.0802	$Y_{min} = -0.025$
	3	0.0	0.0	0.0681	$X_{min} = 0.009$
	4	0.02	-0.02	0.0629	Reg. = 0.0569
	5	-0.02	0.0	0.1236	Acc. = 0.0508
					Move / Use 2R mesh
2	6	-0.02	0.02	0.0599	Reg. Eq. 6.7
	7	-0.01	0.01	0.0531	$Y_{min} = -0.026$
	8	-0.02	0.01	0.0506	$X_{min} = 0.011$
	9	-0.03	0.01	0.0505	Reg. = 0.0502
	C	-0.02	0.0	0.0629	Acc. = 0.0497
					Move / Use 4R mesh
3	10	-0.025	0.015	0.0510	Reg. Eq. 6.8
	C	-0.02	0.01	0.0506	$Y_{min} = -0.025$
	11	-0.025	0.01	0.0502	$X_{min} = 0.012$
	C	-0.03	0.01	0.0505	Reg. = 0.0502
	13	-0.025	0.005	0.0540	Acc. = 0.0495
					Converge / Stop

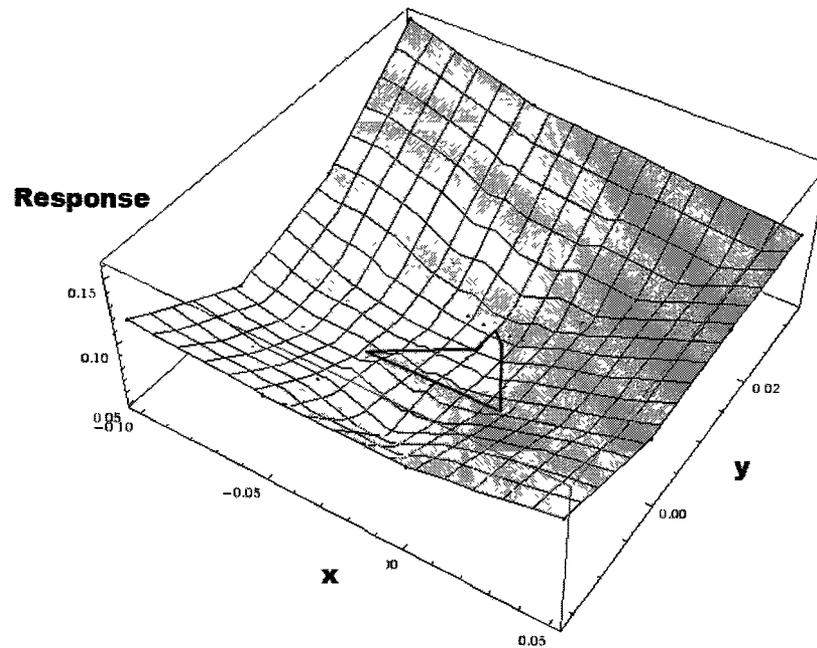


Figure 6.6: The least-square direct-search algorithm results (Table 6.4) are illustrated in blue. It is compared to the original direct-search in red (Table 6.3) to show the path followed by either algorithms to the minimum.

direct-search algorithm does 6 iterations in comparison to the least-square direct-search with 3 iterations. Even the first iteration in the least-square algorithm gives a very close estimate to the minimum in contrast to the original direct-search algorithm in which the first iteration is far from the final results. In terms of CPU time, each thermal analysis took about 3 minutes and each stress analysis took about 45 minutes (total 48 minutes). The L_9 Taguchi DOE matrix has 9 tests run on 3 cores with total of 2 hours and 24 minutes. The η -and-R DOE matrix has 36 tests run on 4 cores totalling 7 hours and 12 minutes. Table 6.5 and 6.6 summarized the computation time and the number of cores used in each iteration for the original direct-search DOE matrix and the least-square direct-search DOE matrix respectively.

Table 6.5: Total CPU time and the number of cores used for each iteration in the original direct-search DOE matrix.

Iteration	Num. of Cores	Num. of Analyses	CPU Time
1	3	9	2 h 24 m
2	3	3	48 m
3	4	8	1 h 36 m
4	4	4	48 m
5	4	4	48 m
6	4	4	48 m
		32	Total 7 h 12 m

Table 6.6: Total CPU time and the number of cores used for each iteration in the least-square direct-search DOE matrix.

Iteration	Num. of Cores	Num. of Analyses	CPU Time
1	3	5	1 h 36 m
Min test	1	1	48 m
2	4	4	48 m
Min test	1	1	48 m
3	3	3	48 m
Min test	1	1	48 m
		15	Total 5 h 36 m

6.8 CWM Results for the Optimum

The objective of the optimization problem was to minimize the final deflection. Fig. 6.7 shows the final deflection for these cases: the weld with no side heater, with the side heater only but without a weld, and the weld with the side heater. Distance is from the left bottom corner to the right bottom corner of the bar for deflection in [m].

Residual stress provides key information for designers along with the information on deflection. Fig. 6.8 shows the longitudinal residual stress in the bar after welding is complete for the weld with no mitigation, when the side heater only is applied and for the weld mitigated by the side heater. Residual stress is plotted for a line normal to the weld, Y-direction, from the top edge to the bottom edge of the bar at the mid-length of the the bar. Units are [Pa] and [m] for stress and distance respectively. The weld with side heater significantly reduces the residual stress in this cross-section.

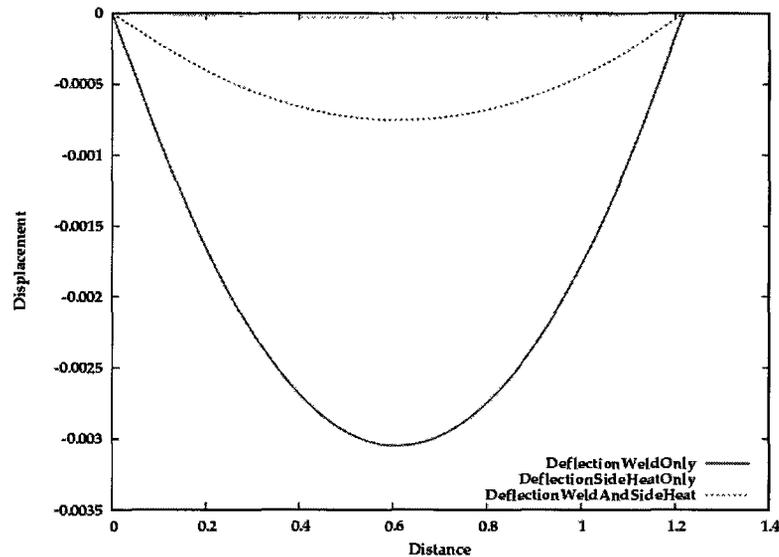


Figure 6.7: Final deflection [m] for the weld with no mitigation, with the side heater only applied, and with the weld mitigated by the side heater. Distance is from the left bottom corner to the right bottom corner of the bar. Units are [m].

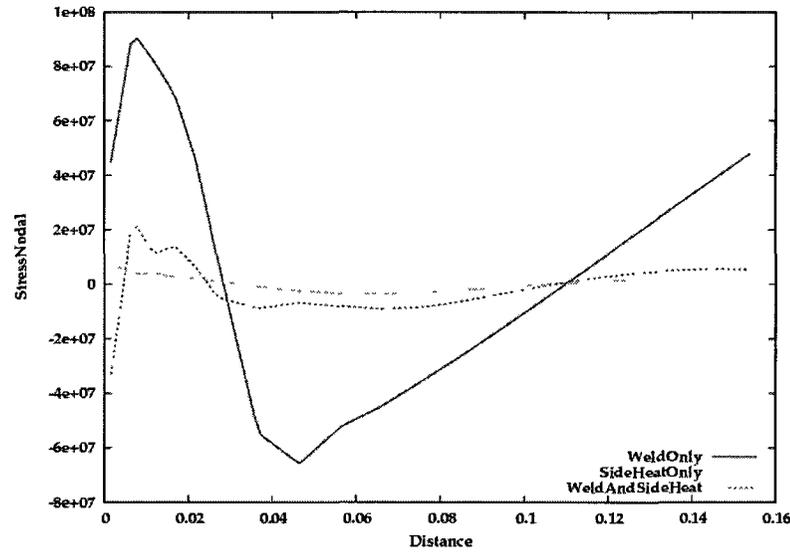


Figure 6.8: Longitudinal residual stress, σ_{xx} , in the bar after welding is complete for the weld with no mitigation, when the side heater only applied and the weld mitigated by the side heater.

6.9 Closure of the Analysis

A least-square direct-search method is presented in this analysis to solve a constrained CWM optimization problem with a sequence of DOE matrices. Its performance is compared to an original direct-search algorithm. The CWM optimization problem is to minimize distortion of an edge-welded bar by using a side-heater technique with four design parameters. The optimal solution is a heat source side heater with 6 cm radius, power of 3588 J, located 2.5 cm behind the weld arc and 1 cm above the longitudinal center line of the bar minimizes the deflection. The deflection is plotted for the optimal design to show the improvement. In addition it is shown that minimizing the distortion, also reduces the longitudinal residual stress significantly. The method could be used to minimize the residual stress by solving a DOE matrix if one chooses an objective function.

Direct-search algorithms are more attractive with DOE matrix analyses that solve tens of computer models to explore or map the associated design space specified by

a DOE matrix to find optimal designs. In a multi-project framework, direct-search algorithms that use DOE matrices for more efficient searching are preferred. Such DOE matrix search takes advantage the fact that multiple trial solutions can be obtained simultaneously in contrast to sequential optimization that does a search based on one-result at a time.

Coupling a least-square approximation in a regular direct-search algorithms followed the path to the minimum more efficiently in the neighborhood of a smooth basin. However the least-square approximation is not expected to work well when the response surface is not a smooth basin, e.g., if the response surface is very wavy or rough.

Chapter 7

A Continuous Design Space Exploration; Regression Modelling

Project: Analysis of Predicted Residual Stress in a Weld and Comparison with Experimental Data Using Regression Model

Two uncertainty analyses are conducted to investigate how variation in welding parameters contributes to the performance, i.e., residual stress. The analysis considers, first, a single parameter using an accurate model and, second, four highly interrelated parameters using a regression approximation model for perturbation analysis since a required sample space of 4 parameters using the accurate model is not affordable. Uncertainty analyses are based on the Monte Carlo method and the idea is that if a high number of tests for an accurate model is not feasible, then extracting and using a regression approximation model from the accurate model is an effective.

7.1 Analysis Background

Frequently, the manufacturing process generates residual stress in the structure before in-service loading and this residual stress changes the in-service behavior of structure.

Welding often adds a considerable local residual stress to a welded structure. Residual stresses are caused by incompatible internal strains and they vary over the life cycle of a structure. The residual stresses are usually not known to design engineers in many cases and lack of such knowledge leads to conservative design practices that are expected to increase costs and reduce the performance of welded structures.

Anna Paradowska et al. [38] embark on a research program to measure experimentally residual stresses in welds. The measurement can be done by neutron/X-ray diffraction based on Bragg's reflection equation. However it is expensive to accurately measure residual stresses. The high cost of such measurements further limits the use of experimental measurements by designers to control and mitigate the effect of residual stresses. Besides, the measurements are inherently discrete and often limited to a few points. This small set of data points is a coarse sampling of a full 3D field of the residual stress in a welded structure.

Prediction of residual stresses by numerical modeling is a desirable complement to experimentally measured residual stresses. Although recent numerical modeling is a powerful tool for predicting residual stress, validation with a reference to experimental results is essential to have confidence in the model and the algorithm [22]. Validation is not complete unless it considers the variation induced by uncertainties in parameter values. For example, material properties or process parameters for numerical modeling often are seen as a fixed value in computation although they are not fixed value in reality and their estimation involves inevitable inherent variation.

Designers would like to manage the undesirable effects of residual stresses in as-welded structures by an appropriate selection of material, welding process, structural geometry and fabrication sequence parameters. This can be done best if the designer has a reliable predictive model that can be used in the early design stages. A full 3D field of the residual stress in a welded structure can be computed for cost of one run of the predictive model providing that the model is validated.

Uncertainty analysis uses a set of sample points taken from variations in control parameters. The size of such sample space defines the cost of running the uncertainty analysis. In practical weld simulation, the number of uncertain control parameters results in a large sample space that is often not affordable using an accurate model. One alternative, for such cases, is to use an approximation model constructed on accurate results of a course grid points of the sample space.

7.2 Experimental Data

Anna Paradowska et al. have presented a set of experimental data for residual stress [38] measured by neutron diffraction. The specimen was a bead-on-plate weld and residual stress was measured in the unrestrained specimen after the plate cooled to room temperature.

Their experimental set up for the weld is shown in Fig. 7.1. The specimen is a $100 \times 200 \times 12$ [mm] low-carbon steel plate . The filler metal is 14 [mm] wide and 6 [mm] high. The composition of the plate and filler metal are given in Table 7.1. The flux-cored arc welding process used a 1.6 [mm] diameter electrode with a 20 [mm] contact tip to work distance. ARGOSHIELD 52 shielding gas was used with a gas flow rate of 18 [ℓ /min]. The welding parameters were 260-280 [ampers], 28-30 [volts], and a welding speed of 6 [mm/s]. During welding the specimen rested on a plate but was free to move with no constraints other than rigid body motion.

Table 7.1: Chemical composition of parent and consumable materials [wt%]

Composition	C	Mn	Si	S	P	Ni	Cr	Mo	Cu	V	Ti	Co	Al
Parent metal	0.12	0.63	0.13	0.01	0.01	0.02	0.02	0.01	0.01	<0.01	<0.01	<0.01	0.03
Weld metal	0.10	1.76	0.68	0.02	0.02	0.05	0.03	0.04	-	0.04	-	-	-

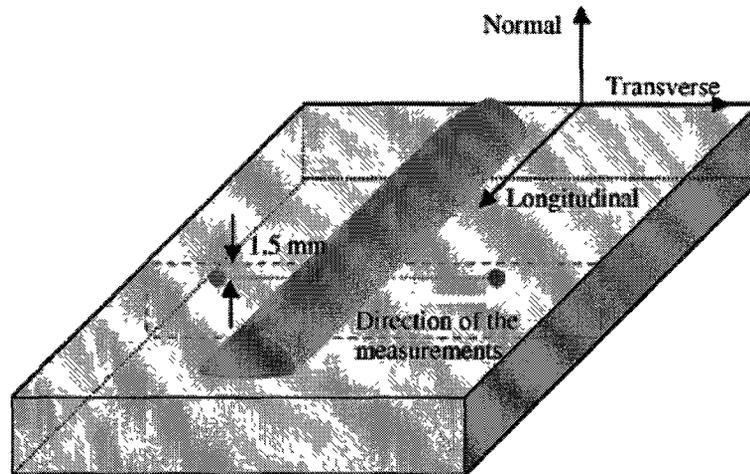


Figure 7.1: Experimental setup that shows the filler metal on the 200x100x12 [mm] plate and the line along which the elastic strain profile is measured

7.3 Computational Setups

A three-dimensional model and meshes were created to accurately model the 3D experimental set up described in section 7.2. The mesh had 3628 8-node brick and 166 6-node prism elements including 5724 nodes. The computing time on a single core of a 2.66 GHz Intel Core 2 quad computer was 39 seconds for thermal analysis, 18 seconds for microstructure analysis and 12.2 minutes for stress analysis.

The thermal analysis of the weld used a double ellipsoid model for welding arc. The dimensions of the double ellipsoid were width $b=7$ [mm], height $c=6$ [mm], front length $a_2=7$ [mm], and trailing length $a_1=4$ [mm]. The transient analysis used 51 time steps including a cool-down period with exponentially increasing time step sizes using coefficient of 1.2.

Fig. 7.2 illustrates the snapshots of the transient temperature fields at times 3.5 seconds and 25 seconds.

Fig. 7.3 and 7.4 illustrate the result of a microstructure analysis including alpha or ferrite, pearlite, martensite, and hardness distribution.

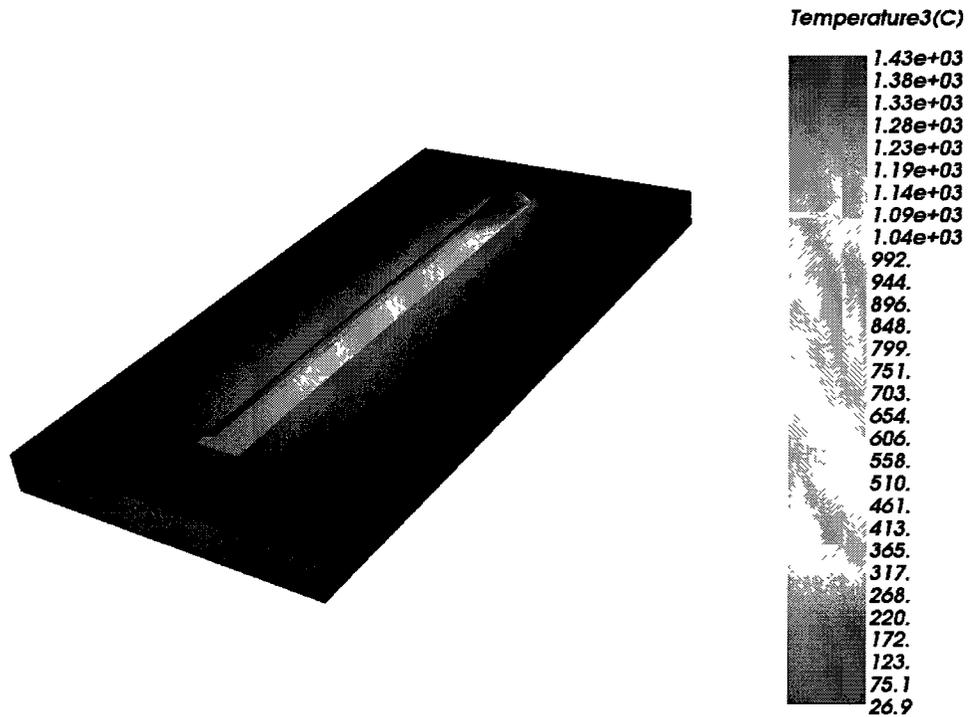
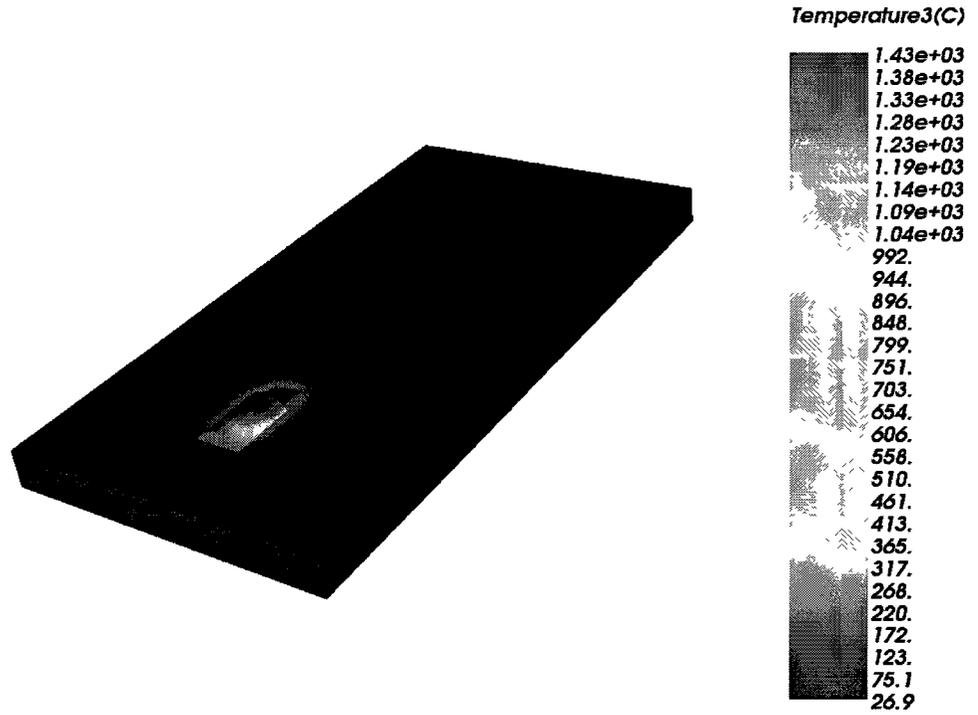


Figure 7.2: Thermal results showing transient temperatures with filler metal being added, after 3.5 seconds, top, and 25 seconds, bottom.

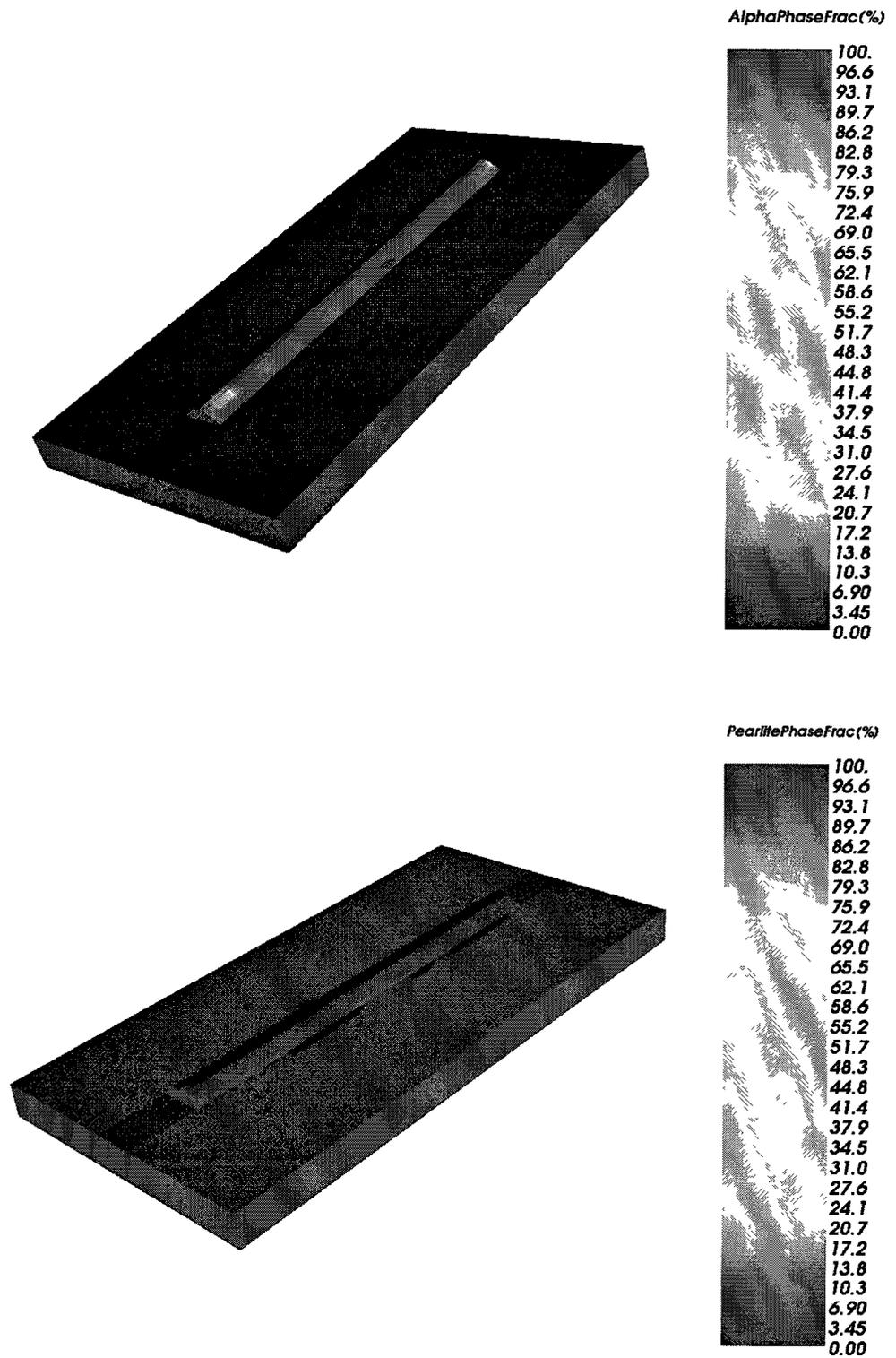


Figure 7.3: Microstructure analysis including alpha or ferrite phase, top, and pearlite phase, bottom.

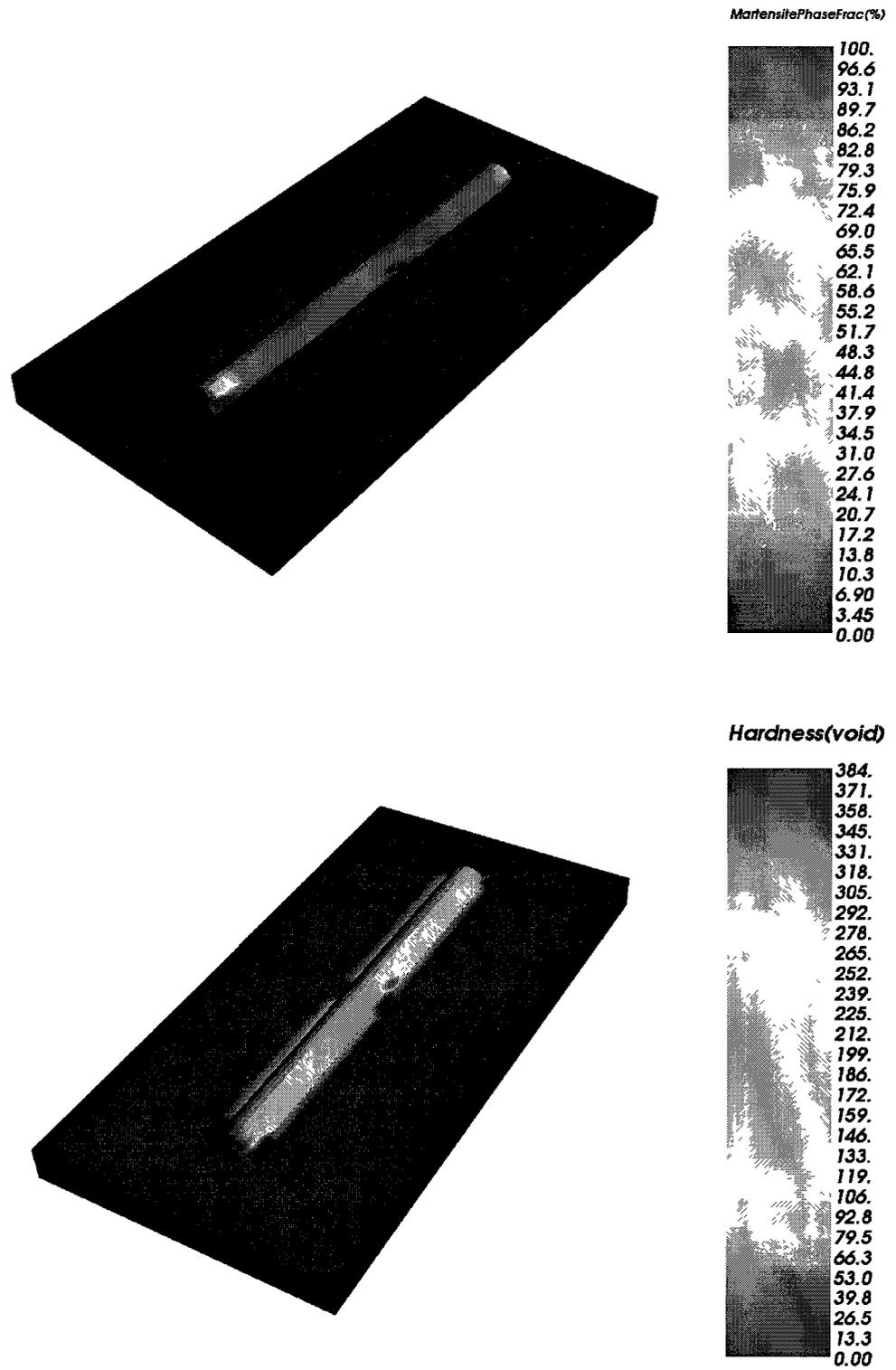


Figure 7.4: Microstructure analysis including martensite phase, top, and hardness distribution, bottom.

Fig. 7.5 shows the total displacement (left) and effective stress (right) in the part after welding, cool down to room temperature and then releasing the restraints.

7.4 Comparison

The same geometry, welding parameters and material data reported in [38] were employed in the analysis. The material property data not provided in [38] were taken from [39] for a similar low-carbon steel with decreasing the room temperature yield stress from 375 [MPa] to 250 [MPa] to agree with the yield stress reported in [38]. The compositions of the base plate and the filler metal are shown in Table 7.1.

The computational results are six components of the stress tensor at each Gauss point and interpolation gives the value of other points. The normal components of nodal stress / residual stress is plotted on the same path of the experimental measurement for comparison.

The residual stress profile for longitudinal, transverse and normal stress components on the same path of the experimental measurement are shown in Fig. 7.6. The solid and dotted lines designate computed and experimental results. Red, black and light blue show the computed longitudinal, transverse and normal components, while dark blue, pink and green show the experimental ones respectively.

Deviation of computed results from experimental data is defined by root mean squared error (RMSE), and the difference between the computed and measured values at each node. This difference is shown in Fig. 7.7. The RMSE is 79, 27 and 58 [MPa] for longitudinal, transverse, and normal components of stress respectively.

To rigorously compare the model to experimental data, one would need replicate sets of experimental data to establish an estimate of the variability required for statistical inference. The accuracy for stress measurement comes from the fact argued in [40] that seeking accuracy of stress values with errors less than $\pm 2\%$ is not realistic.

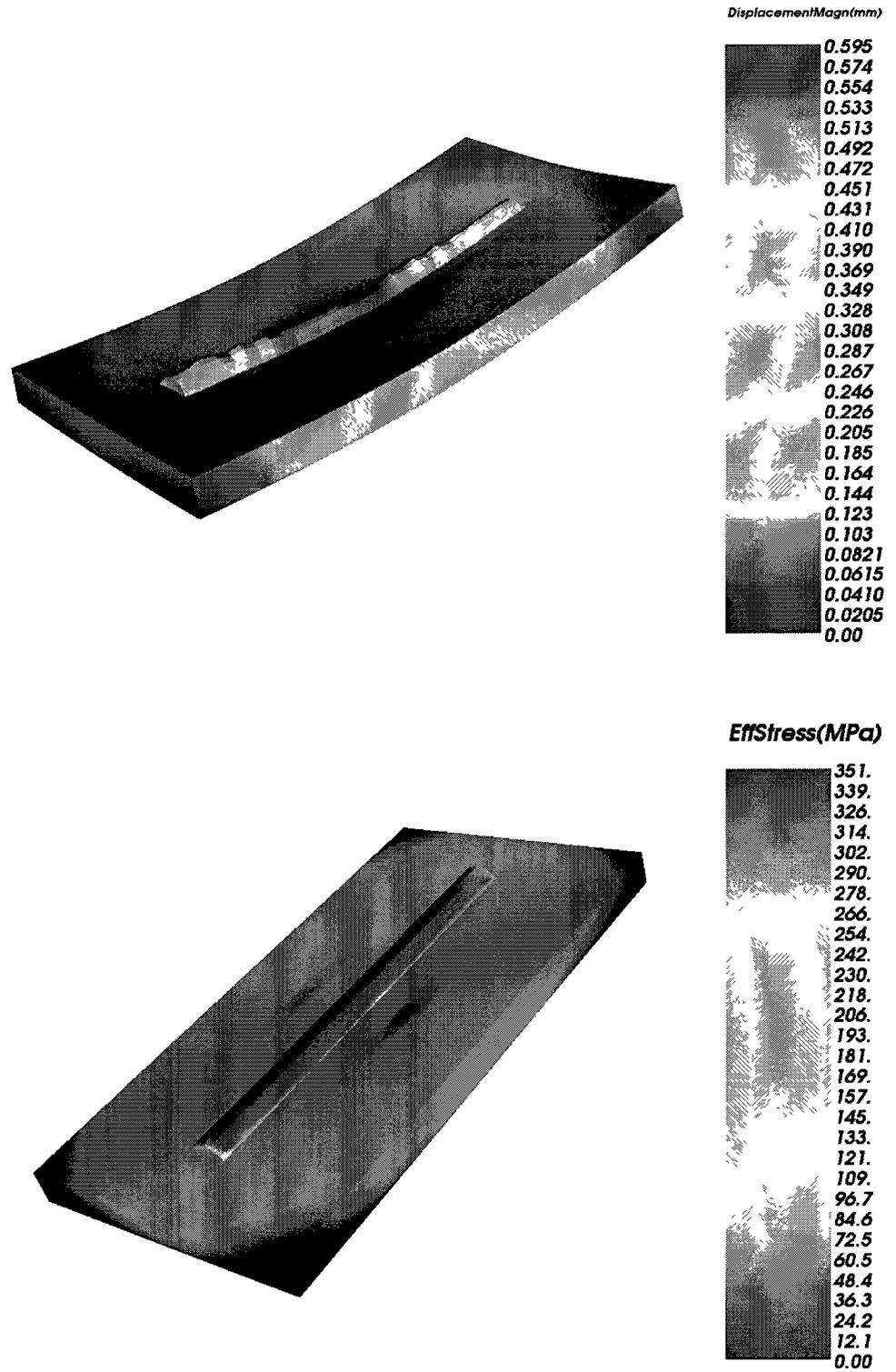


Figure 7.5: Total displacement (left) and effective stress (right) in the part after welding, cool down to room temperature and then releasing the restraints.

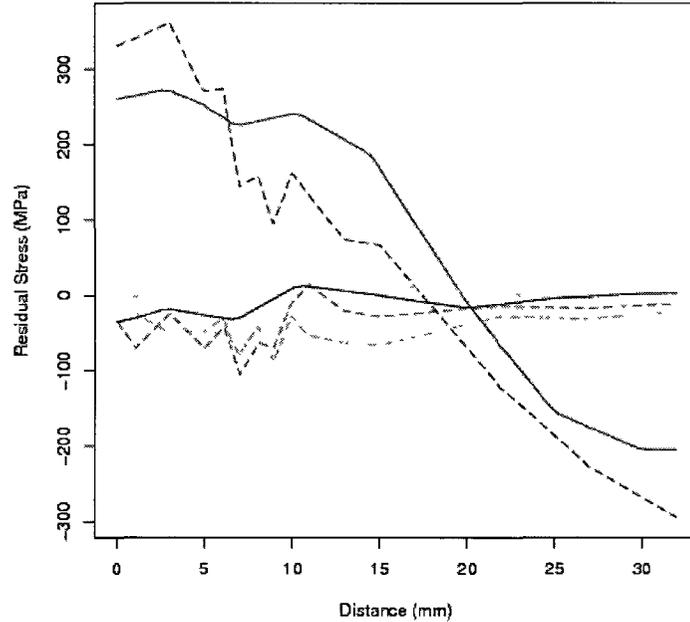


Figure 7.6: Residual stress profile for longitudinal, transverse and normal stress components (see Fig. 7.1) on the same path of the experimental measurement are shown in Fig. 7.1. The solid and dotted lines designate computed and experimental results. Colors; red, black and light blue show the computed longitudinal, transverse and normal components and dark blue, pink and green show the experimental ones respectively.

In [41] Iron and Ahmed argued that in engineering practice an accuracy of $\pm 10\%$ for stress would be considered more realistic. Jonas Pavason et al. [42] conducted Net Benchmark that is a series of tests for three components of residual stress on welding computed by different simulation models including VrSuite and compared with experimental measurements done by different methods. They showed a large scatter in the residual stress measurement. They observed the computational residual stresses lie in the scatted measurements. The confidence in the model is based on the fact

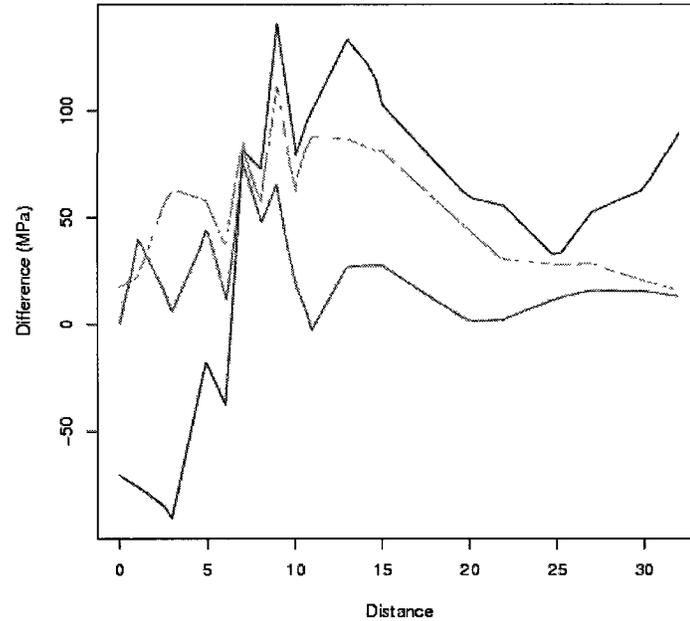


Figure 7.7: The difference between the computed and measured values at each node. Blue is longitudinal, pink is transverse and green is normal stress.

that the model uses the best available known relevant physics. This known physics is based on experiments other than welding. However the model does seem to capture the trend observed in the experiment suggesting that the model would be useful for further sensitivity analysis. Any algorithm that can predict response better can be replaced with no effect in the whole work.

7.5 Uncertainty Analysis

Many uncertainty analysis methods try to capture how uncertain control parameters affect the performance. Monte Carlo method can reliably determine uncertainty by

evaluating a deterministic model using sets of random values for the control parameters as inputs. Reliability of Monte Carlo method results in a frequent usage when the model is complex, nonlinear, or involves many parameters. Monte Carlo method can typically involve thousands of evaluations of the model, a task which may only be practical using super computers.

Monte Carlo method picks sample points from an uncertainty distribution of the control parameter. The welding current in the welding procedure is taken as a single uncertain parameters that affects the residual stress. The nominal value used for current is 280 amperes [39]. Reference [43] monitored experimentally the variation of welding current while welding different sets of experiments and provides an estimate of current distribution during the welding process. Under the assumption of a normal distribution, a normal distribution with mean $\mu = 280$ amperes and standard deviation $\sigma = 5$ was assumed for an uncertainty analysis. These values were obtained from the sample mean and standard deviation from data in [43]. A random sample of size 30 was taken from this distribution so that approximate normality is held by the Central Limit Theorem. This forms a DOE matrix in which each row takes 13 minutes to finish and using 4 cores in parallel, the analysis CPU time to finish the DOE matrix is 1 hour and 44 minutes.

Regarding the selected sample size of 30, each FEM node has 30 different values of the residual stress that gives mean and standard deviation for each node, i.e., the residual stress distribution. Each node has its own distribution's characteristic (μ, σ) that shows the variation in the residual stress due to variation in the welding current.

The experiment line has 78 nodes in the FEM mesh and Fig. 7.8 shows the μ and $\pm 3\sigma$ for the nodes along that line. Red, green and blue lines illustrate higher bound $(+3\sigma)$, mean (μ) , and lower bound (-3σ) respectively. This kind of plot shows how uncertainty of a parameter such as welding current distributes in a performance such as residual stress. It can define sensitive areas on the specimen.

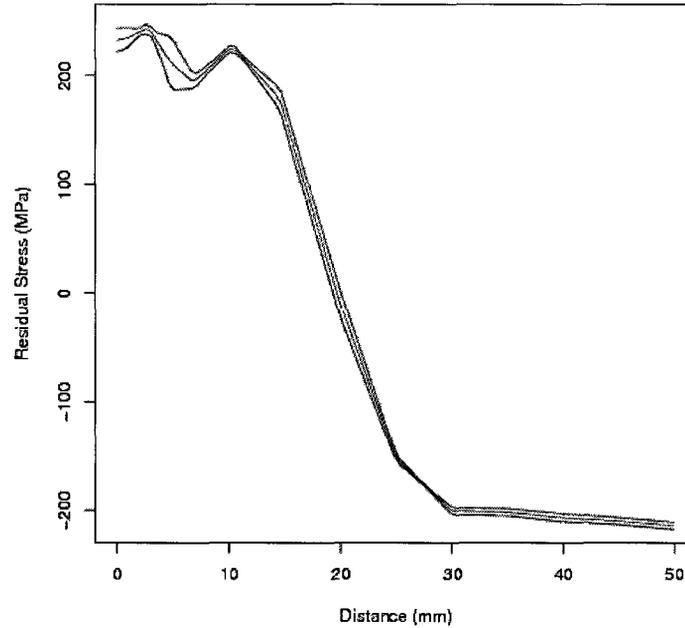


Figure 7.8: Variation in the residual stress due to variation in the welding current for the nodes along the line of experiment. Red, green and blue lines illustrate higher bound, $+3\sigma$, mean, μ , and lower bound, -3σ respectively

This approach is certainly not practical for a large sample space of many control parameters. In this study an alternative way, i.e., point-by-point regression, is employed to accommodate the uncertainty. The idea is to use multidimensional least square regression analysis in which a hyper-plane is fitted to the points positioned in the multidimensional space of the control parameters. The term point-by-point is that each FEM node has its own hyper-plane. The algorithm to implement the point-by-point regression model is presented and used for an uncertainty analysis that captures the residual stress distribution wrt variation of four double-ellipsoid parameters of weld arc.

1. Design of Experiments

A full factorial DOE matrix is employed to give the highest resolution of four parameters with three levels resulting in $3^4 = 81$ rows. The design matrix, X , for the main effects is 81 by 5. Five columns corresponds to *intercept*, a_1 , a_2 , b , and c . Adding two-way interactions and second order of main effects increases the number of columns to 15; $1, a_1, a_2, b, c, a_1a_2, a_1b, a_1c, a_2b, a_2c, bc, a_1^2, a_2^2, b^2, c^2$ and the final design matrix, X , becomes 81 by 15 (Table 7.2). The two-way interactions can be interpreted as the area formed by two lengths. A variation of ± 1 [mm] is applied to each double ellipsoid parameter. The levels were selected to cover a reasonable range of weld pool sizes. However any interval spacing could be utilized. Table 7.3 shows the high, mid, and low values assigned.

Table 7.2: Design matrix, $X_{81 \times 15}$

<i>RunNo.</i>	<i>Constant</i>	a_2	a_1	b	c	a_2a_1	a_2b	a_2c	a_1b	a_1c	bc	a_2^2	a_1^2	b^2	c^2
1	1	6	13	6	4	78	36	24	78	52	24	36	169	36	16
2	1	7	13	6	4	91	42	28	78	52	24	49	169	36	16
3	1	8	13	6	4	104	48	32	78	52	24	64	169	36	16
.
.
.
81	1	8	15	8	6	120	64	48	120	90	48	64	225	64	36

Utilizing 4 cores in parallel, the DOE matrix is finished in 4 hours and 33 minutes including thermal, micro-structure, and stress analysis on a desktop quad core machine.

Table 7.3: Values of each level of double ellipsoid parameters [mm]

Parameter	High [mm]	Mid [mm]	Low [mm]
a_1	15	14	13
a_2	8	7	6
b	8	7	6
c	6	5	4

2. Response Matrix

Each FEM node has a set of 81 values computed from the DOE matrix evaluation. Writing in a matrix form gives responses matrix $Y_{81 \times j}$ where j is the total number of nodes. Back to the measurement line, j is 78 that is the total number of nodes along this line, and the response matrix is $Y_{81 \times 78}$ shown in Table 7.4.

Table 7.4: Response matrix scheme $Y_{81 \times 78}$

	$node_1$	$node_2$	$node_3$. . .	$node_{78}$
run_1	$Y_{1\ 1}$	$Y_{1\ 2}$	$Y_{1\ 3}$. . .	$Y_{1\ 78}$
run_2	$Y_{2\ 1}$	$Y_{2\ 2}$	$Y_{2\ 3}$. . .	$Y_{2\ 78}$
run_3	$Y_{3\ 1}$	$Y_{3\ 2}$	$Y_{3\ 3}$. . .	$Y_{3\ 78}$
.
.
.
run_{81}	$Y_{81\ 1}$	$Y_{81\ 2}$	$Y_{81\ 3}$. . .	$Y_{81\ 78}$

3. Regression Model

The regression model 7.1 is solved by least-squares which has a closed form solution given by equation 7.2 where X and Y are the design and response matrix respectively. Therefore β should be 15×78 (Table 7.5).

$$Y = X\beta + \epsilon \quad (7.1)$$

$$\hat{\beta} = (X^T X)^{-1} X^T Y \quad (7.2)$$

Table 7.5: Coefficient matrix scheme $\hat{\beta}_{15 \times 78}$

	$node_1$	$node_2$	$node_3$.	.	.	$node_{78}$
1	$\beta_{0\ 1}$	$\beta_{0\ 2}$	$\beta_{0\ 3}$.	.	.	$\beta_{0\ 78}$
a_2	$\beta_{1\ 1}$	$\beta_{1\ 2}$	$\beta_{1\ 3}$.	.	.	$\beta_{1\ 78}$
a_1	$\beta_{2\ 1}$	$\beta_{2\ 2}$	$\beta_{2\ 3}$.	.	.	$\beta_{2\ 78}$
b	$\beta_{3\ 1}$	$\beta_{3\ 2}$	$\beta_{3\ 3}$.	.	.	$\beta_{3\ 78}$
c	$\beta_{4\ 1}$	$\beta_{4\ 2}$	$\beta_{4\ 3}$.	.	.	$\beta_{4\ 78}$
$a_2 a_1$
$a_2 b$
$a_2 c$
$a_1 b$
$a_1 c$
bc
a_2^2
a_1^2
b^2
c^2	$\beta_{14\ 1}$	$\beta_{14\ 2}$	$\beta_{14\ 3}$.	.	.	$\beta_{14\ 78}$

The dimensions of $\hat{\beta}$ show the number of terms in the design matrix (columns in Table 7.2) and the number of FEM nodes selected (columns in Table 7.4). $\hat{\beta}$ gives a set of coefficients for a regression function that belongs to each node to approximate the residual stress for that node. Eq. 7.3 gives the approximation, \hat{Y} , for given a_2 , a_1 , b , and c parameters.

$$\widehat{Y}_{1 \times 78} = X_{1 \times 15} \widehat{\beta}_{15 \times 78} \quad (7.3)$$

Given a_1, a_2, b, c in the range of Table 7.3, form X which is a row of $[1, a_1, a_2, b, c, a_1 a_2, a_1 b, a_1 c, a_2 b, a_2 c, bc, a_1^2, a_2^2, b^2, c^2]$, and Eq. 7.3 approximates the response, \widehat{Y} for the given a_1, a_2, b, c . This shows the values of residual stress approximated for the sample nodes.

To show the accuracy of the approximation, the residual stress from arbitrary values of $a_1 = 15, a_2 = 6, b = 7, c = 4$ is compared between FEM and the point-by-point regression model as illustrated in Fig. 7.9. It shows clearly how well this regression model emulates an FEM prediction.

The point-by-point regression can be done for total FEM nodes in the domain and can approximate any user-defined response from FEM. This can be implemented for every time step of the FEM analysis.

7.6 Perturbation Analysis

Equation 7.3 is a continuous well-conditioned mathematical equation. The computing cost depends on the dimension of the matrices but is expected to be significantly cheaper than the FEM. Therefore it is well suited for perturbation analysis such as Monte Carlo method with an large numbers of sample points.

A normal distribution is employed for the four double ellipsoid parameters. This seems a reasonable assumption however other distributions could be used based on the application.

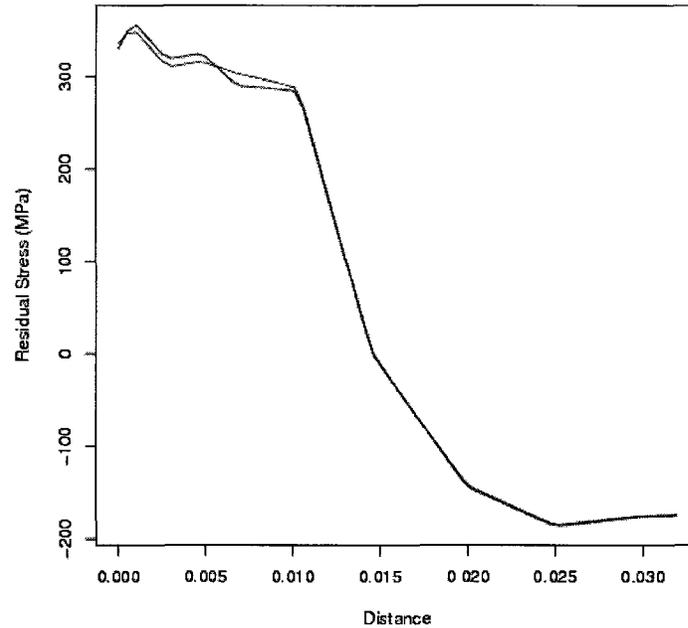


Figure 7.9: Comparing FEM and the point-by-point regression for longitudinal component of residual stress along the measurement line. Red line is FEM prediction and green line shows regression approximation.

Ten thousand (10,000) random sample points are generated for the double ellipsoid parameters and the residual stress is computed along the measurement line using the point-by-point regression model. The CPU time was about 5 minutes for 10,000 evaluations. Fig. 7.10 illustrates the uncertainty distribution using the regression model from a normal variation of the double ellipsoid parameters. The mean value is plotted in green and shows the average residual stress. The maximum and minimum occurrence among the 10,000 sampling are plotted in red and purple respectively. The blue line represents the 95.4 percentile and brown the 4.6 percentile.

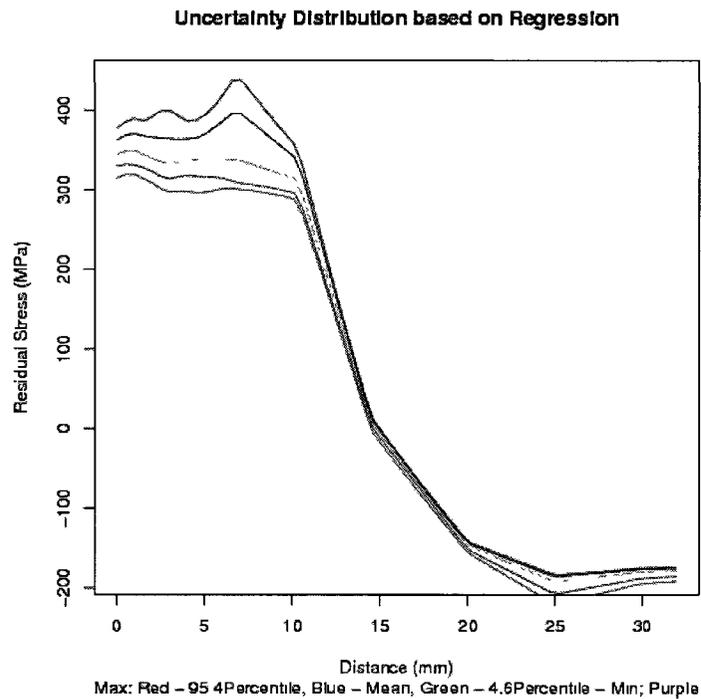


Figure 7.10: Uncertainty distribution using regression model for a normal variation of double ellipsoid parameters.

In fact, this plot is extracted from a histogram of 78 nodes along the measurement line showing the distribution of residual stress. For example, Fig. 7.11 shows this distribution for a node distanced 7.5 millimeter from the weld center line. This is the point with a significant variation from the mean. The histogram in Fig. 7.11 shows that for a normally distributed parameters, there is a very rare chance of getting a residual stress 400 [MPa], i.e, about 70 [MPa] more than nominal value. The 95 % and 99 % percentiles for this histogram are 315378 and 308396 [MPa] respectively.

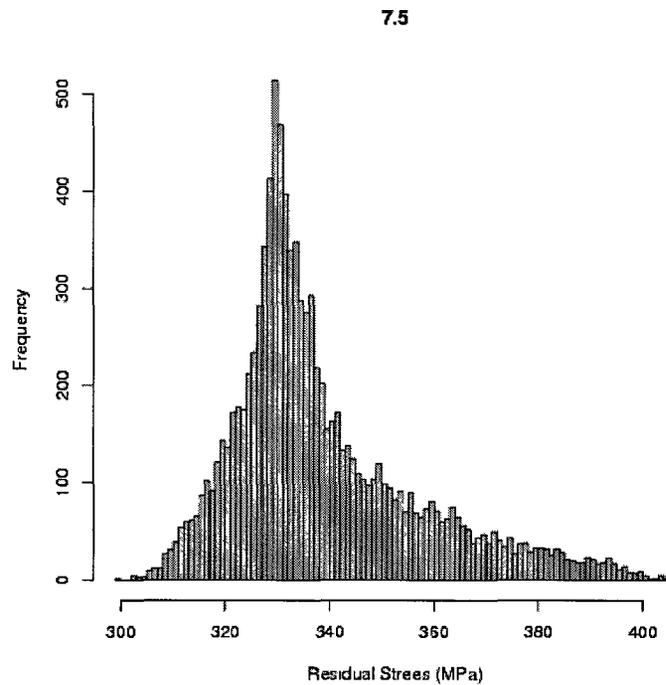


Figure 7.11: Distribution of residual stress for the node distanced 7.5 millimeter from the middle of the weld path

7.7 Closure of the Analysis

Numerical models are well developed now to predict the behavior of many systems. However uncertainty analyses using the numerical models often require a large number of evaluations that makes it unaffordable. Although, many effective stochastic algorithms offer a methodology to accommodate the uncertainty in parameters, they are limited by the computational power available to a certain usually small number of parameters. One alternative is to use approximation models. Regression models are well enough understood to provide a good approximation of the system based on the response of the system. However these models can be valid for only a specific

range of each parameter and this range is usually small particularly when the number of parameters is large. The idea of breaking the whole range of parameters into a combination of pieces with small range and extracting a different regression model for each one, can be a good mathematical solution but it is limited by the cost and time of running a different test for each range and extracting its specific model. Combining regression modeling and numerical modeling may develop a solution to reduce the limitations of both sides. The idea is to employ numerical modeling such as FEM to provide a nodal response field and then create a node-by-node regression model from FEM analysis to cover the whole field. This approximation model can be used for a large sample space such as Monte Carlo method uncertainty analysis. This argument is true if two prerequisites are satisfied. Firstly, the numerical model can capture the behavior of the interest in an acceptable band compared to the experimental data. Secondly, the pieces of range selected for the regression model satisfies the statistical confidence level (adaptive refinement might be needed).

Chapter 8

A Continuous Design Space Exploration; Control and Sensitivity Analysis

Project: Sensitivity of Hot Cracking to Weld Procedure Parameters and Transverse Traction

The section focuses on hot cracking susceptibility analysis. A post-processor was first developed and used in a computational weld mechanics (CWM) framework to predict the transient 3D hot cracking susceptible region for welded structure. A stress approach was used for the hot cracking analysis previously analyzed by Zacharia in [44]. The specimen was a $50 \times 25 \times 0.25$ [mm] 316 stainless steel sheet welded by TIG process with a constant transverse force applied on the side surfaces. The tensile traction, welding power, welding speed, and 4 double ellipsoid shape parameters are varied for a sensitivity analysis of hot cracking wrt the transverse tensile tractions and several welding speeds for which power per unit length is kept constant. A control problem solved to adjust 4 double ellipsoid shape parameters for different welding speeds in constant power per unit length. The analysis includes 117 analyses for the control part and 28 analyses for the sensitivity part both of which are implemented in an automated mode of the CWM framework.

8.1 Analysis Background

In a homogeneous solidification of most materials, the solidification occurs between the liquidus and solidus temperatures. In the late stage of this range, a small fraction of liquid phase remains between the boundaries of the solid phase. This causes a low ductility or brittleness in the material and a tensile stress and/or strain can crack the material. Such a situation may occur likely in welding. The hot cracking requires 4 conditions to occur; 1) temperature in cooling thermal profile; 2) temperature in the Brittleness Temperature Range (BTR) range; 3) strain rate greater than a certain rate to initiate the crack, i.e., CST; and 4) tensile stress / strain to grow the crack. Fig. 8.1 presents the hot cracking susceptibility region. The temperature between the solidus and liquidus is considered as BTR, and CST (Critical Strain Rate) is the critical strain rate such that the hot cracking is susceptible for higher rates. This can be related to the standard definition of strain rate as Eq. 8.1 [45] and [46].

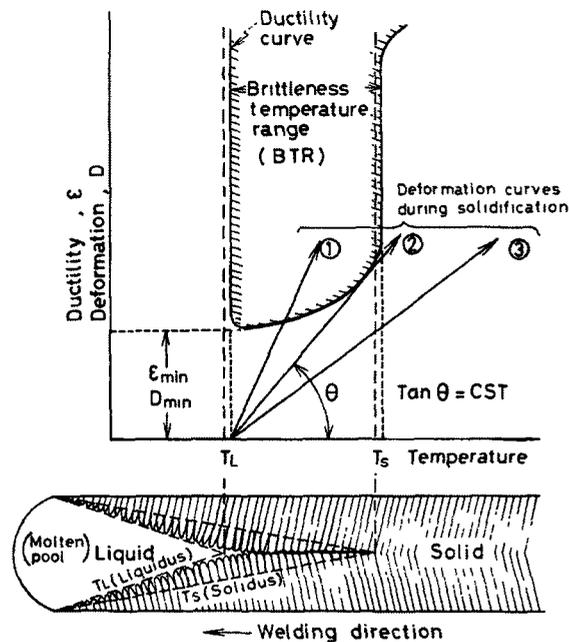


Figure 8.1: Strain approach to define the hot cracking susceptibility region.

$$\frac{d\varepsilon}{dT} = \frac{d\varepsilon/dt}{dT/dt} > CST \quad (8.1)$$

The solidus line cannot characterize a material's brittleness temperature range in actual applications. The BTR can be defined experimentally by techniques such as MISO presented by Matsuda et al. to plot ductility curves in [47] and [48]. Fig 8.2 shows the ductility curves for some stainless steels taken from [48].

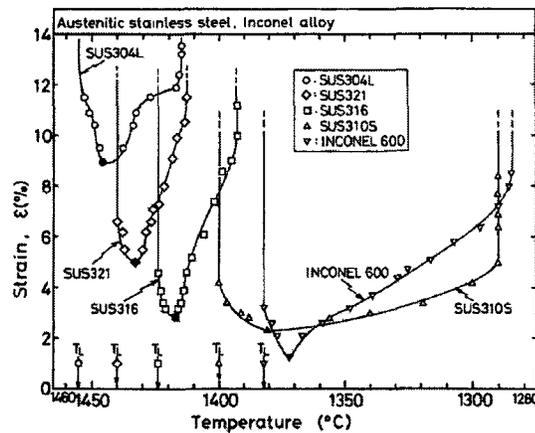


Figure 8.2: Ductility curves for some stainless steels [48].

Zacharia [44] used the tensile stress envelope and the BTR to define the hot crack susceptibility region. He assumed an intersection between tensile envelope and the BTR can cause a hot crack as presented in Fig. 8.3.

A post-processor is developed and used to check the 4 conditions for hot cracking occurrence over all gauss point results to find the intersection. The intersection shows the hot crack susceptibility region which could, in fact, be empty.

8.2 Computational Setup

The full computational model that includes thermal and stress analysis is analyzed by VrWeld software [21]. The details of the model setup for transient thermal and

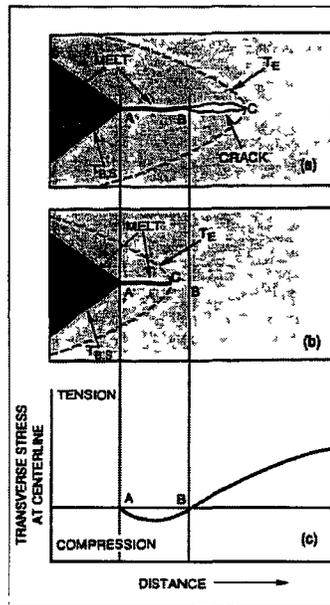


Figure 8.3: Stress approach to define hot cracking susceptibility region.

stress analysis are described below. The test is taken from [44] that is a $50 \times 25 \times 0.25$ [mm] 316 stainless steel sheet metal with chemical composition C 0.018, Ni 12.16, Cr 17.04, Mo 1.98, Mn 1.70, Si 0.34, Cu 0.05, N 0.047, P 0.032, S 0.010 [Wt %]. There is a jig that applies a transverse stress which is modeled numerically as pressure boundary conditions (BCs) on side edges. The mesh employed shown in Fig. 8.4 has 4,600 8-node brick elements and 7,171 nodes. There is one element through the thickness and the weld path is in the mid-width along the Z from the Z-negative to Z-positive edge.

Temperature dependent material properties were employed in the analysis. The gas tungsten arc weld was made along the specimen centerline. The welding parameters are given in Table 8.1. The operating voltage was not in [44] and the voltage of 10 [volts] gives a closer weld width (around 2 [mm]). The analysis in [44] was done until the weld traveled 30 [mm] from the starting edge and the results were reported for the weld pool at this location of weld. Therefore this analysis stops when the weld reaches the same location. The total analysis has 61 time steps of welding with no

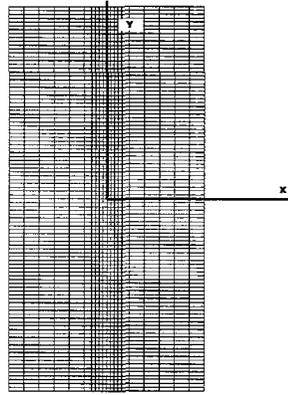


Figure 8.4: Bottom view the mesh employed in this analysis.

cooling, similar to the test in [44].

Table 8.1: Process Parameters

Welding Parameters Used	
Electrode diameter	Tungsten 1.6 [mm]
Travel speed	14.8 [mm/s]
Arc current	20 [A]
Arc voltage	10
Arc length	0.76 [mm]
Shielding gas	Ar
Flow rate	142 [cm^3/s]

8.2.1 Thermal Analysis Setup

The power density distribution function Q [w/m^3], the ‘Double Ellipsoid’ heat source model [23], was used with arc efficiency 0.8, and double ellipsoid shape parameters; 1, 4, 1,1 [mm] for front, rear, width, depth.

A convection boundary condition generated a boundary flux q [w/m^2] on all external surfaces. This uses ambient temperature of $T_{\text{ambient}} = 300$ [$^{\circ}K$]. The convection coefficient as a function of temperature is given in Eq. A.5.

The time step length while welding was chosen so that in two time steps the heat source was required to travel one element along the weld path. The CPU time on a single core 3.3 GHz Intel was 37 seconds for the thermal analysis.

Fig. 8.5 compares the thermal result taken from [44] and computed here respectively.

8.2.2 Stress Analysis Setup

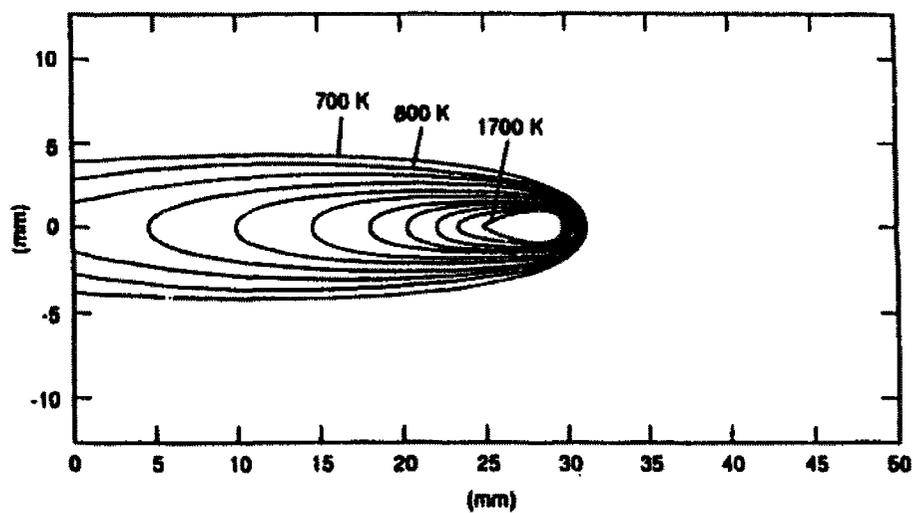
The XYZ directions are defined as transverse, normal and longitudinal to the weld and boundary conditions are rigid body fixities and pressure on the XY side surfaces. To prevent rigid body modes, XY mid-plane (on weld path) was fixed by zero X Dirichlet displacement on 3 nodes, weld path line by zero X and Y Dirichlet displacement, and the weld start point on the weld path by zero X, Y and Z displacement.

Zacharia [44] applied 172 [MPa] stress on the side surfaces and the stress distribution from his work shown in Fig. 8.6. This is a snapshot taken for the time that the weld is at 30 [mm] distance from the start. Fig. 8.7 shows the computed stress distribution from the current setup when a 172 [MPa] stress has been applied. Fig. 8.8 shows the 0 and 310 [MPa] stress contours and Fig. 8.9 shows the 0 and 160 [MPa] stress contours for comparison with Fig. 8.6.

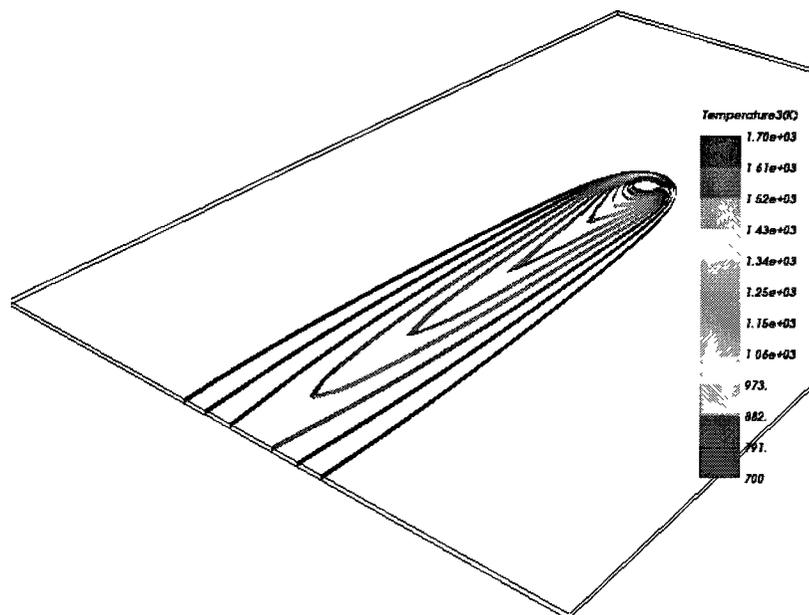
The CPU time on a single core 3.3 Ghz Intel was 5 minutes for the stress analysis.

8.3 Transverse Traction Sampling

Zacharia [44] applied a tensile traction on side surfaces of the sheet metal as a Neumann BC. This BC was active for the full range of time steps.



(a) Thermal result from [44].



(b) Computed result.

Figure 8.5: Comparing computed thermal results and thermal results from [44].

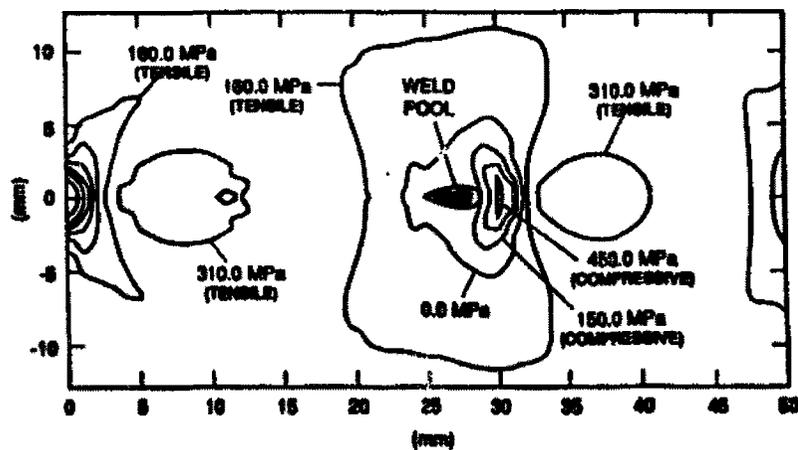


Figure 8.6: Stress distribution from [44]. 172 [MPa] stress is applied on side and weld at 30 [mm] from the start point.

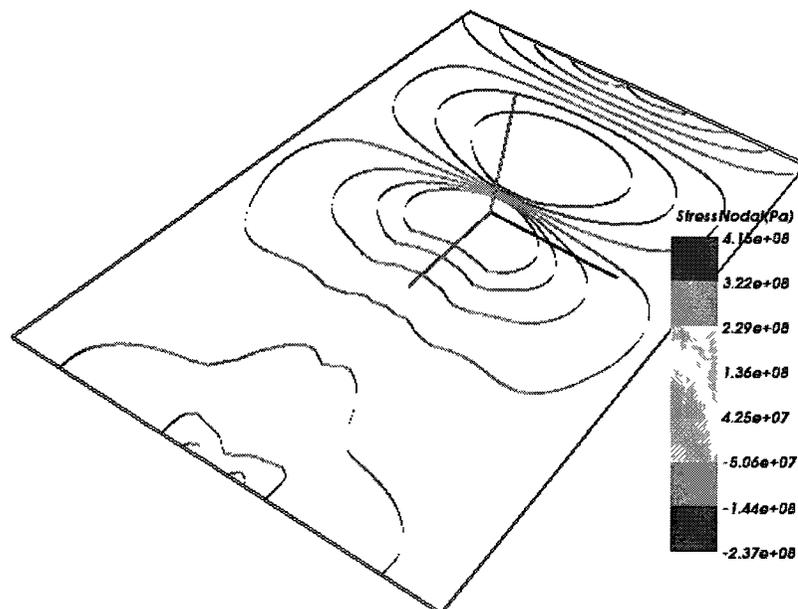


Figure 8.7: Current computed stress distribution. 172 [MPa] stress is applied on sides and weld at 30 [mm] from the start point.

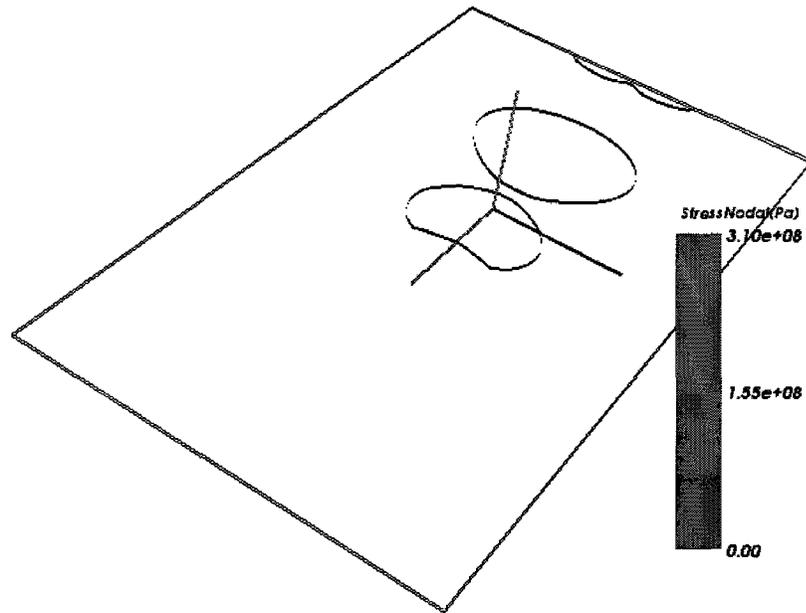


Figure 8.8: 0 and 310 [MPa] stress contours for comparison with Fig. 8.6. 172 [MPa] stress is applied on sides and weld at 30 [mm] from the start point.

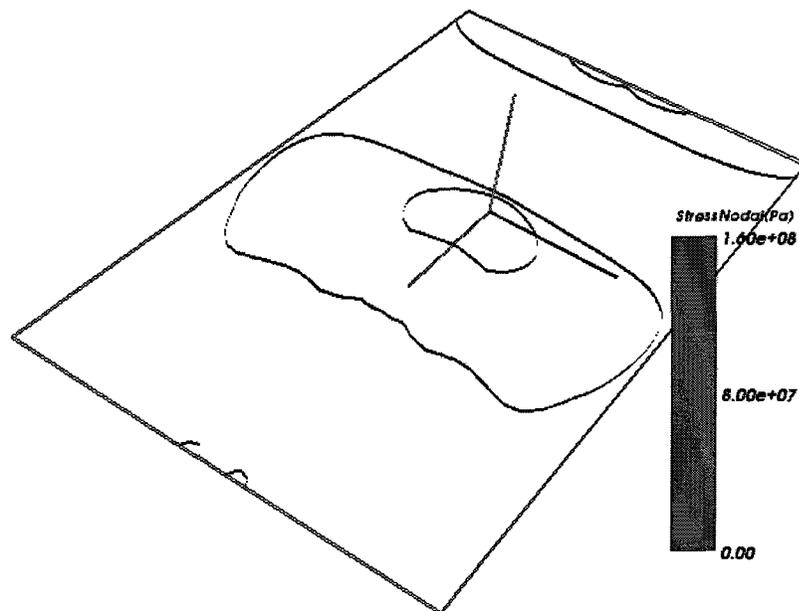


Figure 8.9: 0 and 160 [MPa] stress contours for comparison with Fig. 8.6. 172 [MPa] stress is applied on sides and weld in 30 [mm] from star point.

In the current analysis, four values of traction are sampled to study the effect of a transverse tensile stress on hot cracking. The tractions and corresponding forces applied to the side surfaces are given in Table 8.2.

Table 8.2: Transverse traction and corresponding stresses applied to the side surfaces.

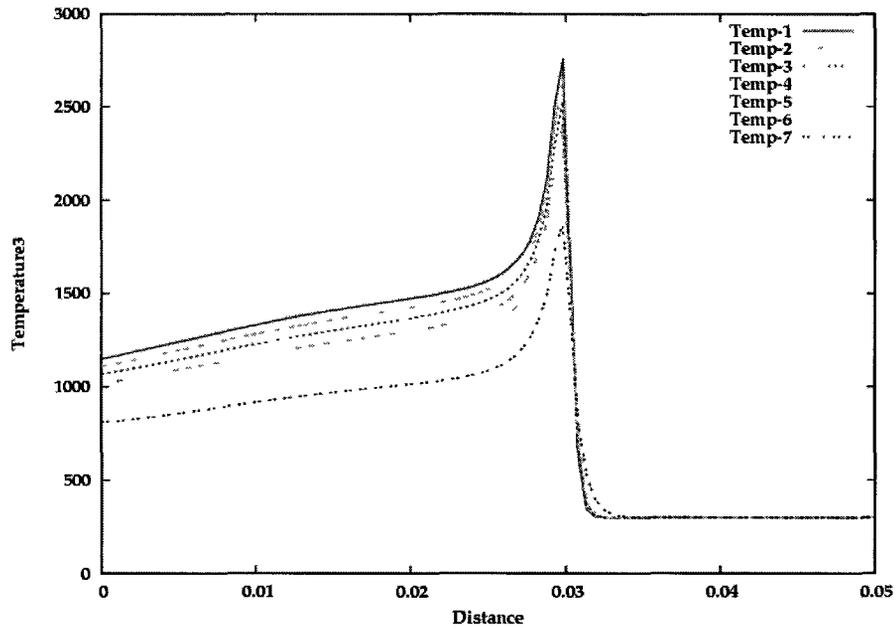
Row	Side Area [mm^2]	Force [N]	Traction [MPa]
1	12.5	0	0
2	12.5	1250	100
3	12.5	1875	150
4	12.5	2500	200

8.4 Power per Unit Length Sampling

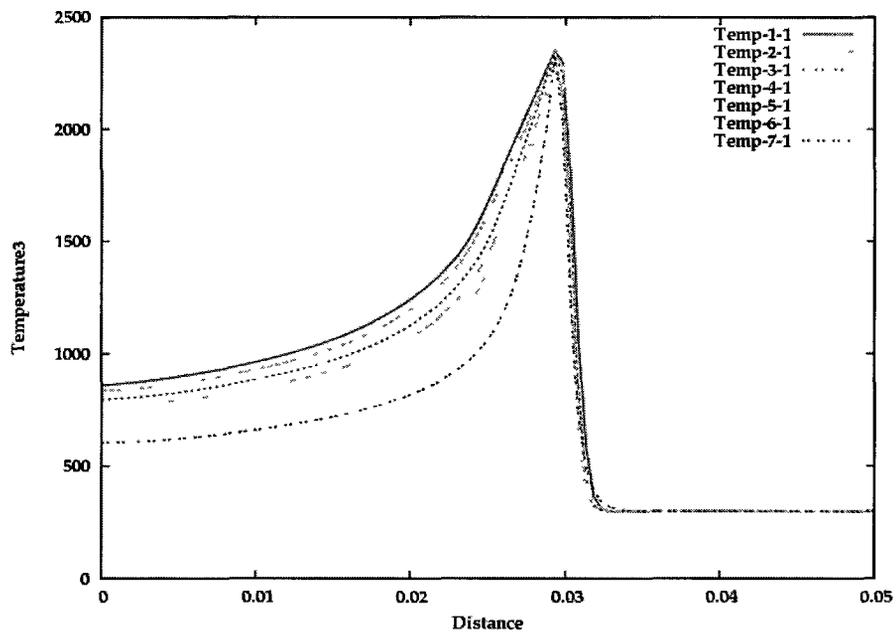
Table 8.3 gives 7 different powers and speeds analyzed for a constant power per unit length. Using these powers and speeds with fixed double ellipsoid parameters generates different maximum temperatures in the weld pool model as shown in Fig. 8.10(a). Therefore the double ellipsoid parameters should be adjusted for a given maximum. A maximum of 2,351 [$^{\circ}K$] that is about 550 above the melting point of stainless steel, and adjusted the double ellipsoid parameters as given in Table 8.4 was assumed. Fig. 8.10(b) shows the thermal profile after the adjustment. The process of adjustment is described in following section.

8.5 Double Ellipsoid Adjustment

The number of FEM nodes in the double ellipsoid can be simplified by picking nodes in a box with dimensions $a_1 + a_2$, b , and c , i.e., double ellipsoid shape parameters.



(a) Before adjustment



(b) After adjustment

Figure 8.10: Thermal Profile for 7 powers and speed given in Table 8.3. X is the distance on the weld path from the start point and the peak shows the weld position.

Table 8.3: Power and speed tested for constant power per unit length.

Row	Power [joule]	Speed [m/s]	Power/Length [j/m]
1	400	0.0296	13,513.5
2	350	0.0259	13,513.5
3	300	0.0222	13,513.5
4	250	0.0185	13,513.5
5	200	0.0148	13,513.5
6	150	0.0111	13,513.5
7	100	0.0074	13,513.5

Table 8.4: Adjusted double ellipsoid parameters.

Row	Power [joule]	Speed [m/s]	Max Temp. [$^{\circ}K$]	a_2 [m]	a_1 [m]	b [m]	c [m]
1	400	0.0296	2346	0.002	0.005	0.0015	0.001
2	350	0.0259	2368	0.0015	0.005	0.0015	0.001
3	300	0.0222	2338	0.0015	0.004	0.0015	0.001
4	250	0.0185	2331	0.001	0.004	0.0015	0.001
5	200	0.0148	2351	0.001	0.004	0.001	0.001
6	150	0.0111	2358	0.0005	0.003	0.001	0.001
7	100	0.0074	2357	0.0002	0.002	0.0005	0.001

This box has more node than real double ellipsoid but does not change the double ellipsoid adjustment because the box and real double ellipsoid are both characterized with the same a_1 , a_2 , b , and c that are adjustment parameters. Since the sheet is thin and there is one element through the thickness, it can be assumed that c , depth, is constant 0.001 [m]. Eq. 8.2 gives the number of nodes, i.e., nodes on top surface of the box, where F, R, D are the number of nodes in front, rear, width of the arc position.

$$n = (F + R + 1)(2D + 1) \quad (8.2)$$

The reference project used $a_2= 0.001$ [m], $a_1= 0.004$ [m], $b= 0.001$ [m], and $c= 0.001$ [m] gives F, R, and D equal to 2, 8, and 2 nodes respectively. Since the element size is 0.005 m around the weld path, one node will be added/removed for every 0.005 m increase/decrease in double ellipsoid shape parameters. Assuming a variation given in Table 8.5 to the references F, R, and D, the DOE-matrix of all combinations for such variation is given in Table 8.6. Row number 14 holds values of the reference project and the total number of nodes is obtain by Eq. 8.2. There exist 13 rows with a greater total number of nodes than the reference denoted by H, and 13 rows with fewer that the total number of nodes than the reference denoted by L.

Table 8.5: Variation for F, R, and D.

Variation	F (a_2)	R (a_1)	D (b)
Increase	3 (0.0015 [m])	10 (0.005 [m])	3 (0.0015 [m])
Reference	2 (0.001 [m])	8 (0.004 [m])	2 (0.001 [m])
Decrease	1 (0.0005 [m])	6 (0.003 [m])	1 (0.0005 [m])

Project number 5 in Table 8.3, i.e., Temp-5 in Fig. 8.10(a), has a peak temperature of 2351 [°K]. Fig. 8.10(a) shows that Temp-1, Temp-2, Temp-3, and Temp-4

Table 8.6: DOE matrix of all combinations for variation given in Table 8.5 to the references F, R, and D.

Row	H/L	F	R	D	Total Nodes, Eq. 8.2
1	H1	3	10	3	98
2	H2	3	10	2	70
3	L1	3	10	1	42
4	H3	3	8	3	84
5	H4	3	8	2	60
6	L2	3	8	1	36
7	H5	3	6	3	70
8	L3	3	6	2	50
9	L4	3	6	1	30
10	H6	2	10	3	91
11	H7	2	10	2	65
12	L5	2	10	1	39
13	H8	2	8	3	77
14	Ref	2	8	2	55
15	L6	2	8	1	33
16	H9	2	6	3	63
17	L7	2	6	2	45
18	L8	2	6	1	27
19	H10	1	10	3	84
20	H11	1	10	2	60
21	L9	1	10	1	36
22	H12	1	8	3	70
23	L10	1	8	2	50
24	L11	1	8	1	30
25	H13	1	6	3	56
26	L12	1	6	2	40
27	L13	1	6	1	24

have a peak temperatures above 2351 [$^{\circ}K$] and Temp-6, Temp-7 have a peak temperatures below 2351 [$^{\circ}K$]. Adding more nodes to the double ellipsoid reduces the peak temperature and removing nodes increases the peak temperature. Rows denoted by H in the DOE matrix given in Table 8.6 add nodes and decrease the peak temperature. For the same reason, rows denoted by L in the DOE matrix given in Table 8.6 remove nodes and increase the peak temperature. Therefore, 13 H-denoted rows applied to projects 1, 2, 3, and 4 to reduce the peak temperature and 13 L-denoted rows applied to the projects 6 and 7 to elevate the peak temperature.

Fig. 8.11, 8.12, 8.13, and 8.14 show the thermal results for the 13 H-denoted rows applied to the projects; 1, 2, 3, and 4 respectively. These thermal profiles are plotted for a transverse, x-direction, line to the weld path from one side of the weld to the other side. The horizontal line is the objective temperature 2351 [$^{\circ}K$]. Fig. 8.11 shows that none of the 13 rows can drop the peak temperature in a certain tolerance of 2351 [$^{\circ}K$] and more nodes need to be added to the double ellipsoid for further decrease in peak temperature. Fig. 8.12 shows that the row H1 is in a certain tolerance of the horizontal 2351 [$^{\circ}K$] line with peak temperature of 2368 and therefore form the decision for the project 2, Table 8.3, allows the choice H1 of Table 8.6. Fig. 8.13 shows that the row H3 is in a certain tolerance of the horizontal 2351 [$^{\circ}K$] line with peak temperature of 2338 and therefore form the decision for the project 3, Table 8.3, allows the choice H3 of Table 8.6. Fig. 8.14 shows that the row H8 is in a certain tolerance of the horizontal 2351 [$^{\circ}K$] line with peak temperature of 2331 and therefore form the decision for the project 4, Table 8.3, allows the choice H8 of Table 8.6.

Fig. 8.15 and 8.16 show the thermal results for the 13 L-denoted rows applied to projects 6 and 7 respectively. Fig. 8.15 shows that the row L12 is in a certain tolerance of the horizontal 2351 [$^{\circ}K$] line with peak temperature of 2358 and therefore form the decision for the project 6, Table 8.3, allows the choice L12 of Table 8.6. Fig.

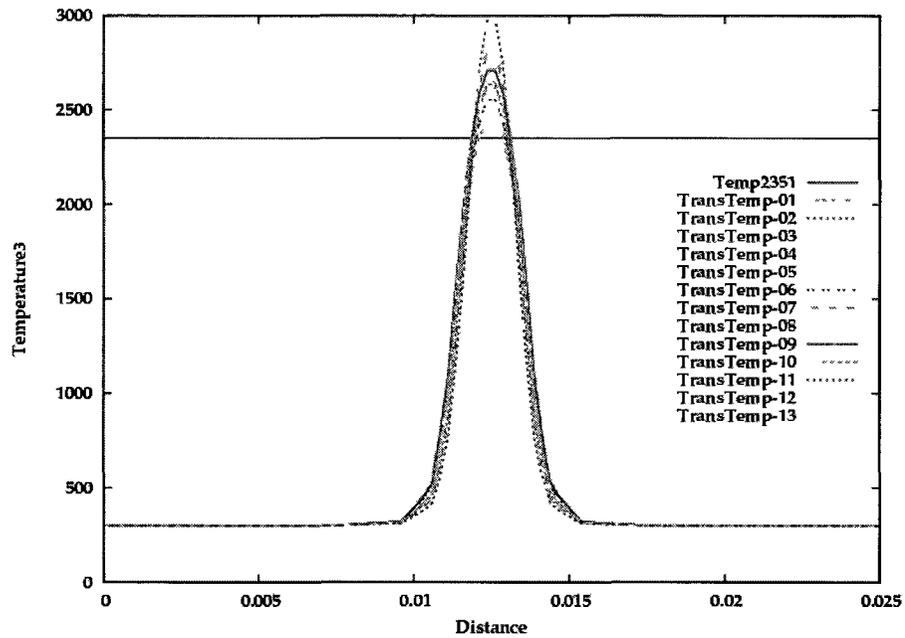


Figure 8.11: Thermal results for the 13 H-denoted rows, Table 8.6, applied to project 1, Table 8.3.

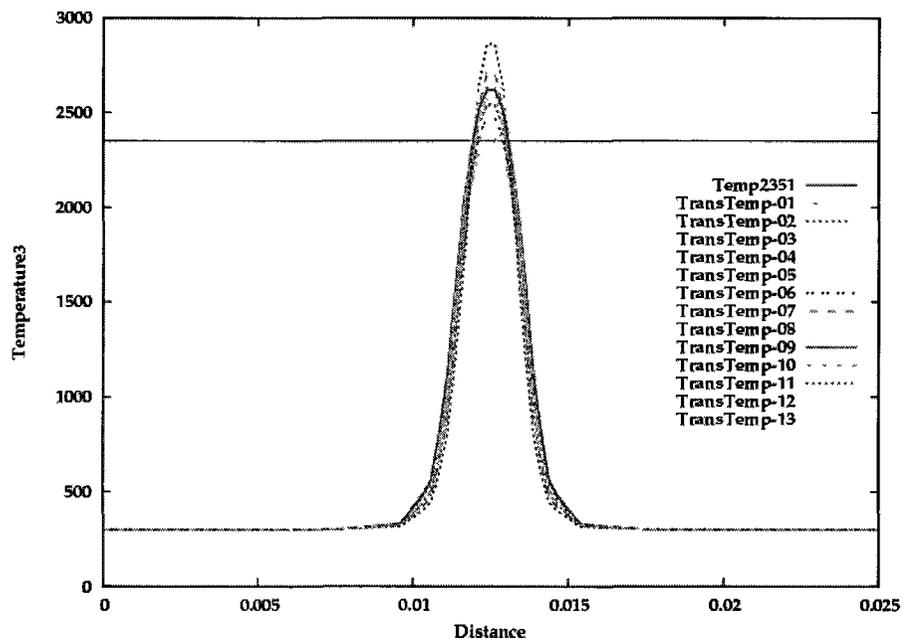


Figure 8.12: Thermal results for the 13 H-denoted rows, Table 8.6, applied to project 2, Table 8.3.

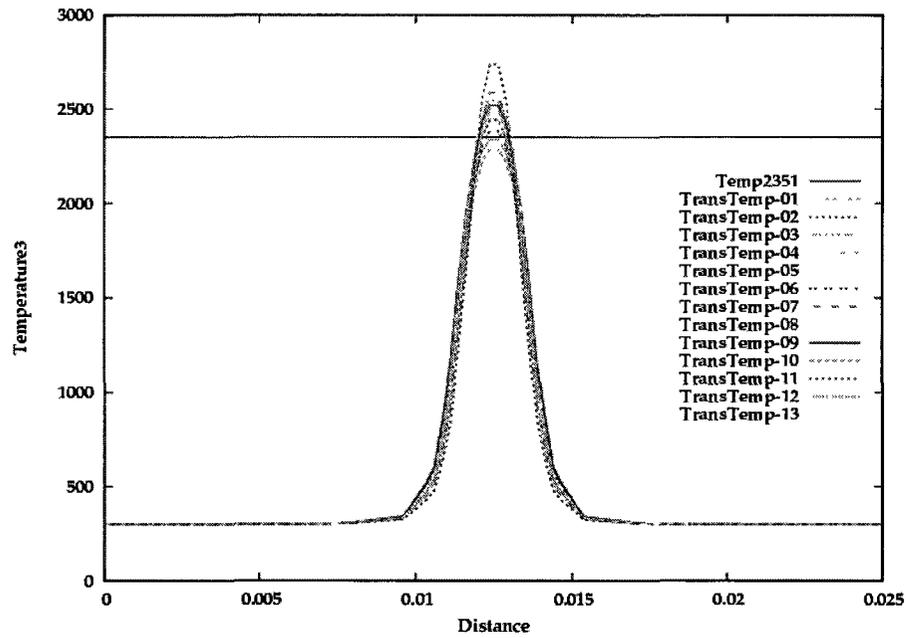


Figure 8.13: Thermal results for the 13 H-denoted rows, Table 8.6, applied to project 3, Table 8.3.

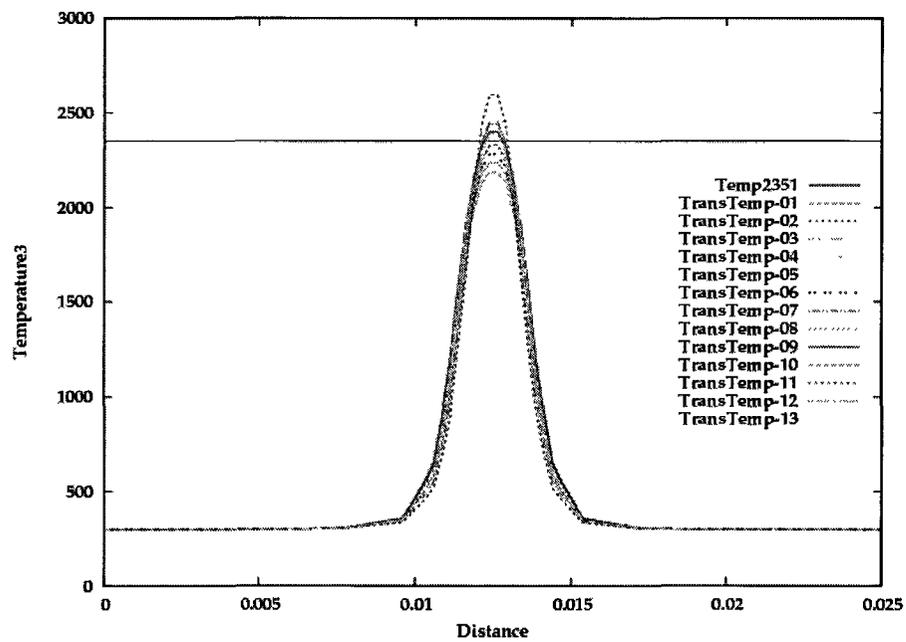


Figure 8.14: Thermal results for the 13 H-denoted rows, Table 8.6, applied to project 4, Table 8.3.

8.16 shows that none of the 13 rows can elevate the peak temperature to a certain tolerance of 2351 [°K] and more nodes need to be removed from the double ellipsoid to further increase the peak temperature.

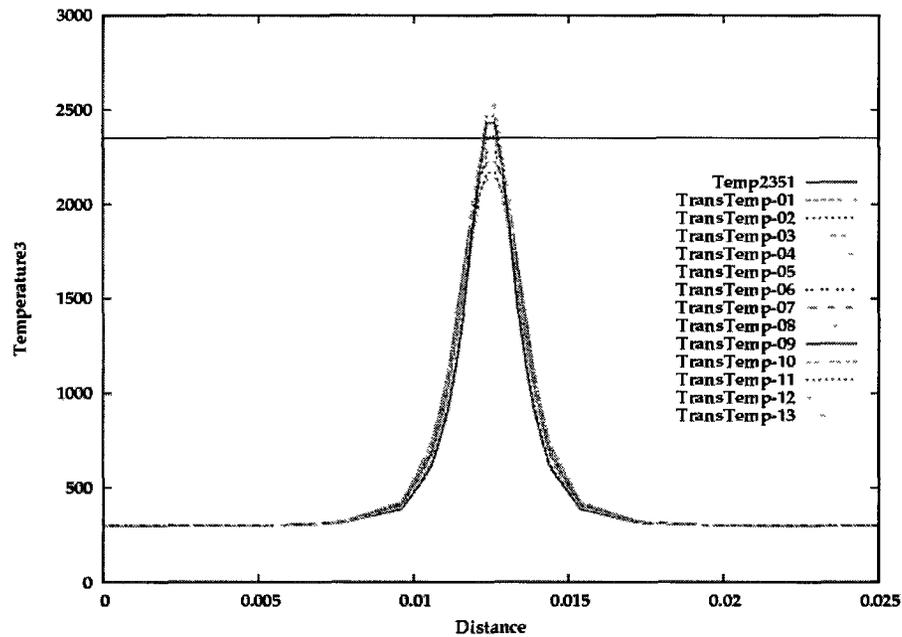


Figure 8.15: Thermal results for the 13 L-denoted rows, Table 8.6, applied to project 6, Table 8.3.

For projects 1 and 7, another iteration of variation in the number of nodes in the double ellipsoid is needed such that more nodes are added to the project 1 and some nodes are removed from project 7. Therefore, the first variation given in Table 8.5, has been updated by Tables 8.7 and Table 8.8 for projects 1 and 7 respectively.

Table 8.7: Second iteration of variation in F, R, and D for project 1.

Variation	F (a_2)	R (a_1)	D (b)
Increase	4 (0.002)	12 (0.006 m)	4 (0.002 m)
Reference	3 (0.0015 m)	10 (0.005 m)	3 (0.0015 m)
Decrease	2 (0.001 m)	8 (0.004 m)	2 (0.001 m)

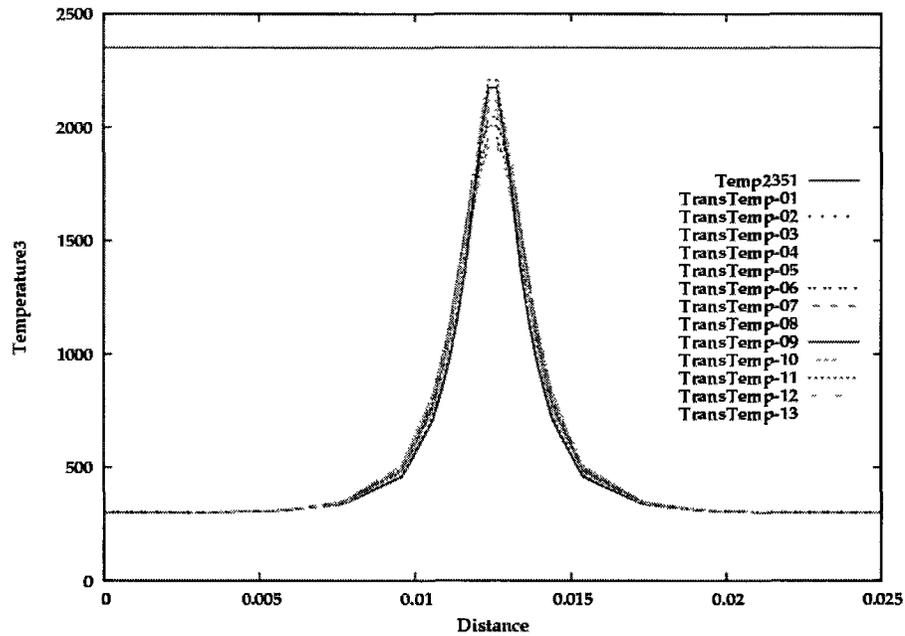


Figure 8.16: Thermal results for the 13 L-denoted rows, Table 8.6, applied to project 7, Table 8.3.

Table 8.8: Second iteration of variation in F, R, and D for project 7.

Variation	F (a_2)	R (a_1)	D (b)
Increase	2 (0.001)	8 (0.004 m)	2 (0.001 m)
Reference	1 (0.0005 m)	6 (0.003 m)	1 (0.0005 m)
Decrease	0 (0.0002 m)	4 (0.002 m)	0 (0.0002 m)

The new variations for project 1 generate the DOE matrix given in Table 8.9. This DOE matrix has 13 rows that are picked from all combinations such that Eq. 8.2 has a value higher than 55, i.e., the reference project. Since the number of nodes could not be zero in one of the dimensions, a minimum size of 0.0002 m is assumed for the double ellipsoid parameters. The new variations for project 7 generate the DOE matrix given in Table 8.10. This DOE matrix has 13 rows that picked from all combinations such that Eq. 8.2 has a value less than 55, i.e., the reference project.

Table 8.9: DOE matrix generated from the second variation given in Table 8.7 for project 1.

Row	F	R	D
H1-1	4	12	4
H1-2	4	12	3
H1-3	4	10	4
H1-4	4	10	3
H1-5	4	8	4
H1-6	3	12	4
H1-7	3	12	3
H1-8	2	10	4
H1-9	3	8	4
H1-10	2	12	4
H1-11	2	12	3
H1-12	2	10	4
H1-13	2	8	4

Fig. 8.17 shows the thermal results for the 13 H1-denoted rows in Table 8.9 applied to the project 1. The row H1-4 is in a certain tolerance of horizontal 2351 [$^{\circ}K$] line with peak temperature of 2346 and therefore the decision for the project 1, Table 8.3, is to pick the double ellipsoid from row H1-4 in Table 8.9.

Table 8.10: DOE matrix generated from the second variation given in Table 8.7 for project 7. '0' denotes the minimum double ellipsoid size of 0.0002 [m].

Row	F	R	D
L7-1	2	8	0
L7-2	2	6	0
L7-3	2	4	1
L7-4	2	4	0
L7-5	1	8	0
L7-6	1	6	0
L7-7	1	4	1
L7-8	1	4	0
L7-9	0	8	0
L7-10	0	6	1
L7-11	0	6	0
L7-12	0	4	1
L7-13	0	4	0

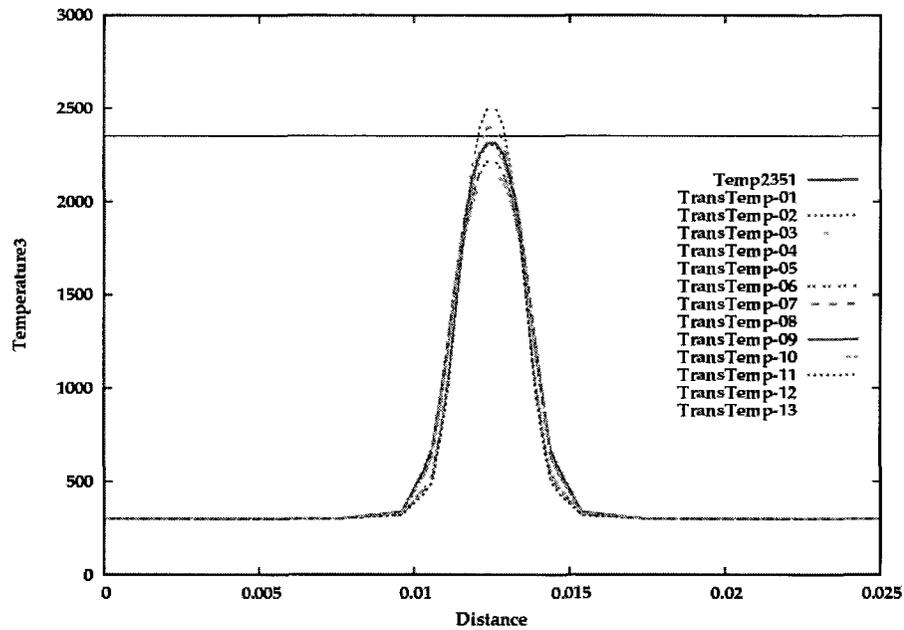


Figure 8.17: Thermal results for the 13 H1-denoted rows, Table 8.9, applied to project 1, Table 8.3.

Fig. 8.18 shows the thermal results for the 13 L7-denoted rows in Table 8.10 applied to the project 7. The rows L7-12 is in a certain tolerance of horizontal 2351 [$^{\circ}K$] line with peak temperature of 2357 and therefore the decision for the project 7, Table 8.3, is to pick the double ellipsoid from row L7-12 in Table 8.10.

The adjusted double ellipsoid size for the 7 powers and speeds of this analysis are given in Table 8.4 and the thermal profile is shown in Fig. 8.10(b). For this adjustment, 9 DOE matrices with 13 projects for each one are implemented. A thermal run on a single core, 3.3 GHz Intel, takes 37 seconds. Each DOE matrix takes 148 seconds using 4 cores. The DOE matrices were implemented by the Multi-project mode of VrWeld that automates the implementation of a DOE matrix as a single run utilizing the available cores. This mode saves user time to setup the projects given in DOE matrices. This represent significant savings. For example, a total of 117 projects has been analyzed for this adjustment. A manual implementation of 117 projects with change in 4 parameters of double ellipsoid size takes weeks for user to

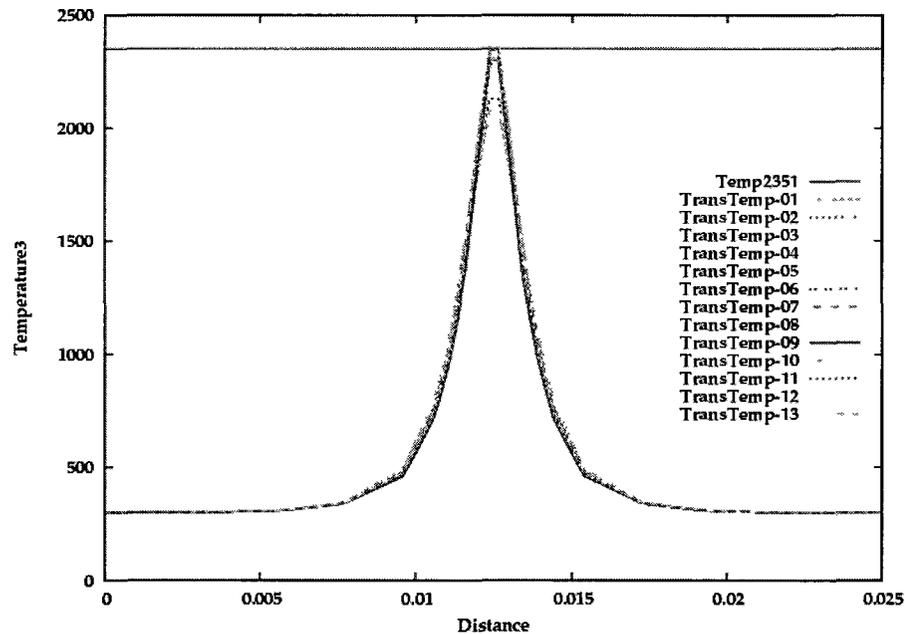


Figure 8.18: Thermal results for the 13 L7-denoted rows, Table 8.10, applied to project 7, Table 8.3.

setup. The automation requires only few minutes of user time for each DOE matrix (not each project). Therefore the user time relates to the design of 9 DOE matrices, i.e., an Excel file with 13 rows and 4 columns, and decision making from the results of each DOE matrix.

8.6 Hot Crack Sensitivity Analyses

A 7D design space is assumed including tensile traction, welding power, welding speed, and 4 double ellipsoid shape parameters. The welding power and welding speed are constrained to keep the power per unit length constant. This results in the power-speed pair satisfying the constraint. For a given power-speed pair, a control problem was solved in the subspace of 4 double ellipsoid shape parameters to keep maximum temperature in the double ellipsoid model in certain level for all pairs. The 7 settings of power, speed, double ellipsoid parameters are given in Table 8.3. These 7 settings

and the 4 values for the tensile tractions, in Table 8.2, are selected to define a sampling grid on the design space. This sample space has 28 design points that come from all combinations between the 4 tractions and 7 settings of speed-power-double-ellipsoid. Writing the 28 analyses with 7 parameters each in the form of a matrix is formally equivalent to a DOE matrix with 28 rows and 7 columns.

Solving 28 settings in such DOE matrix maps the design points to the state space of this problem. A function on the state space which reflects the behavior of interest and is easy to interpret is required. For this hot cracking project, a transverse stress component profile along the weld path could be a good function. Nodes behind the weld pool on weld path are likely candidates for hot cracking because they experience a cooling thermal profile while under tensile stress. Accordingly, the X/transverse component of stress on the weld path from the weld's start point to the other side of the sheet was plotted.

The plot legends have 2 numbers: the first number from left shows the row number from Table 8.3 for one of the 7 settings of speed-power-double-ellipsoid. The second number from left shows the row number from Table 8.2 for one of the 4 tensile traction on the sheet side surfaces.

Fig. 8.19 to 8.22 show the variation in transverse component of stress on the weld path for different speed-power-double-ellipsoid settings when tractions are fixed.

Fig. 8.23 to 8.29 show the variation in transverse component of stress, σ_{xx} , on the weld path for different tractions when speed-power-double-ellipsoid settings are fixed.

8.7 Discussion

The length of the tensile stress region behind the weld on the weld path can characterize the tensile region. Fig. 8.19 to 8.22 show that for given traction, increasing

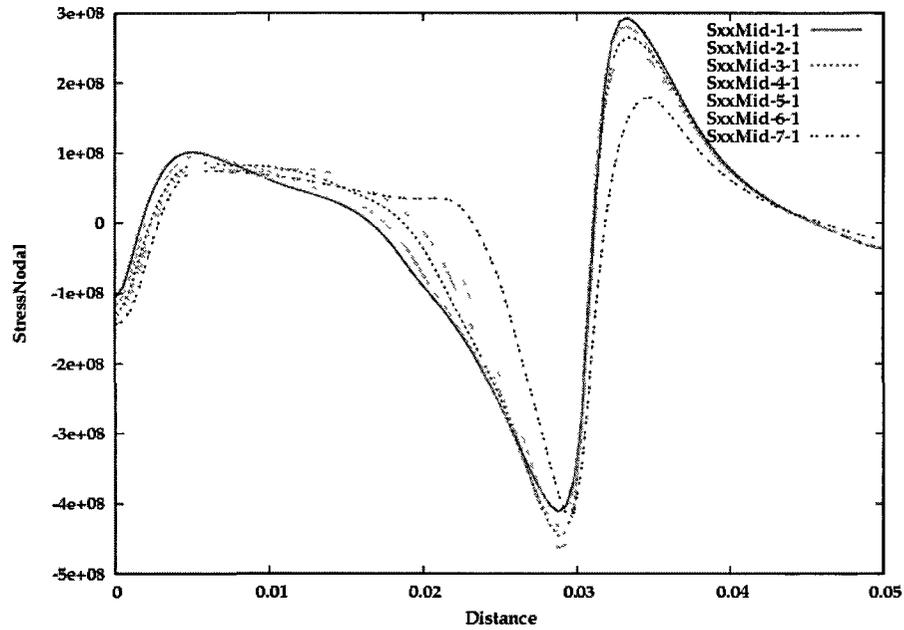


Figure 8.19: Variation in transverse component of stress, σ_{xx} , on the weld path for different speed-power-double-ellipsoid settings when traction is 0 N.

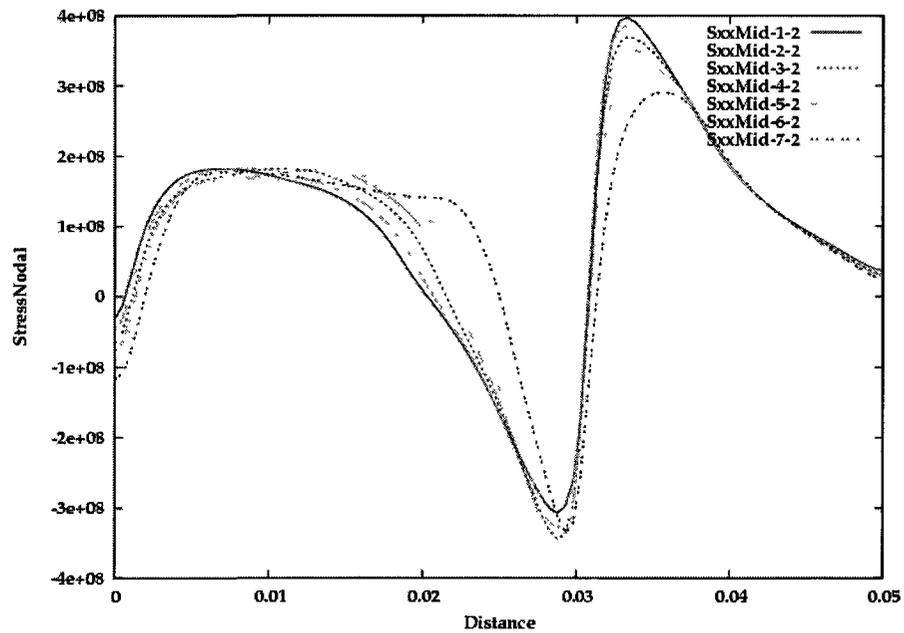


Figure 8.20: Variation in transverse component of stress, σ_{xx} , on the weld path for different speed-power-double-ellipsoid settings when traction is 1250 N.

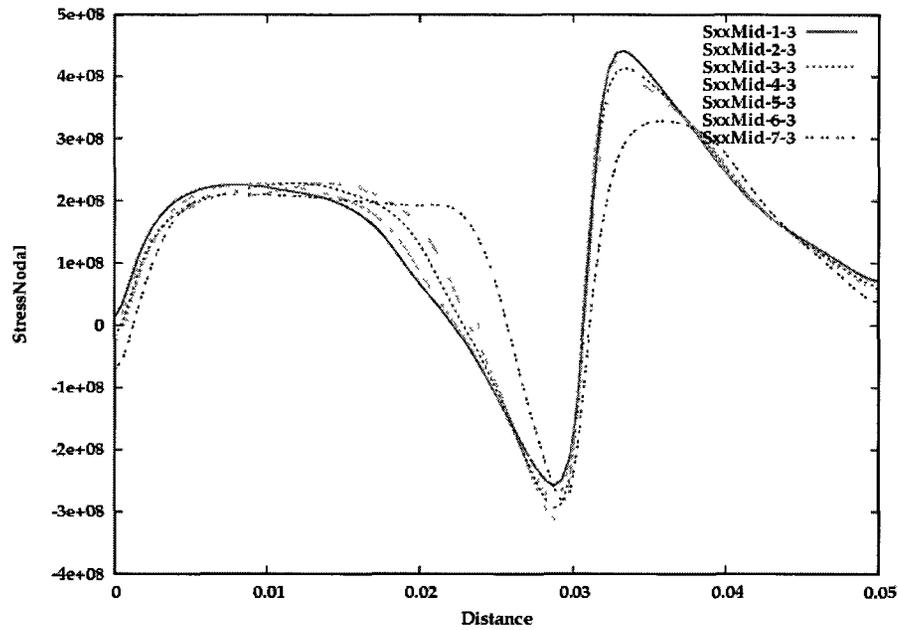


Figure 8.21: Variation in transverse component of stress, σ_{xx} , on the weld path for different speed-power-double-ellipsoid settings when traction is 1875 N.

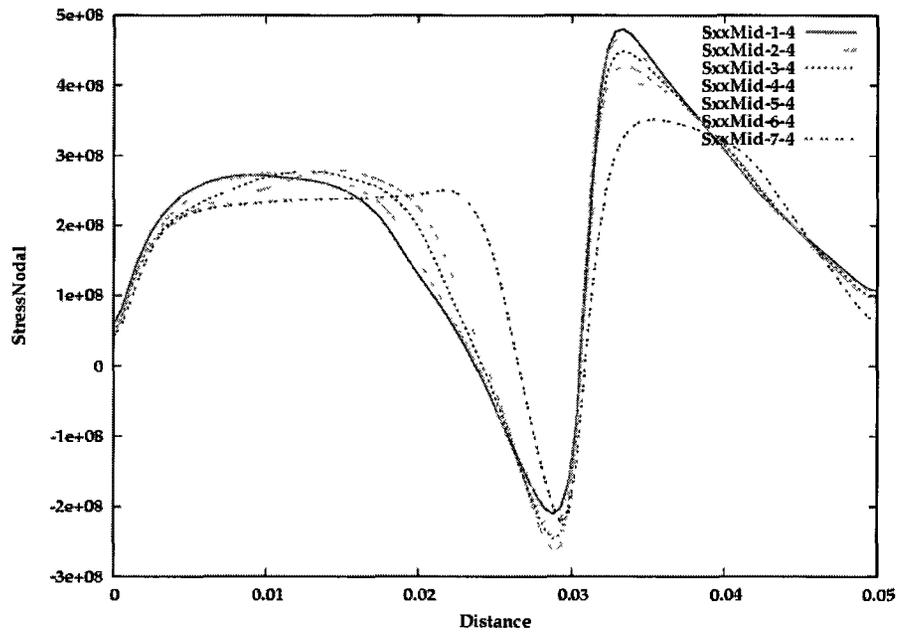


Figure 8.22: Variation in transverse component of stress, σ_{xx} , on the weld path for different speed-power-double-ellipsoid settings when traction is 2500 N.

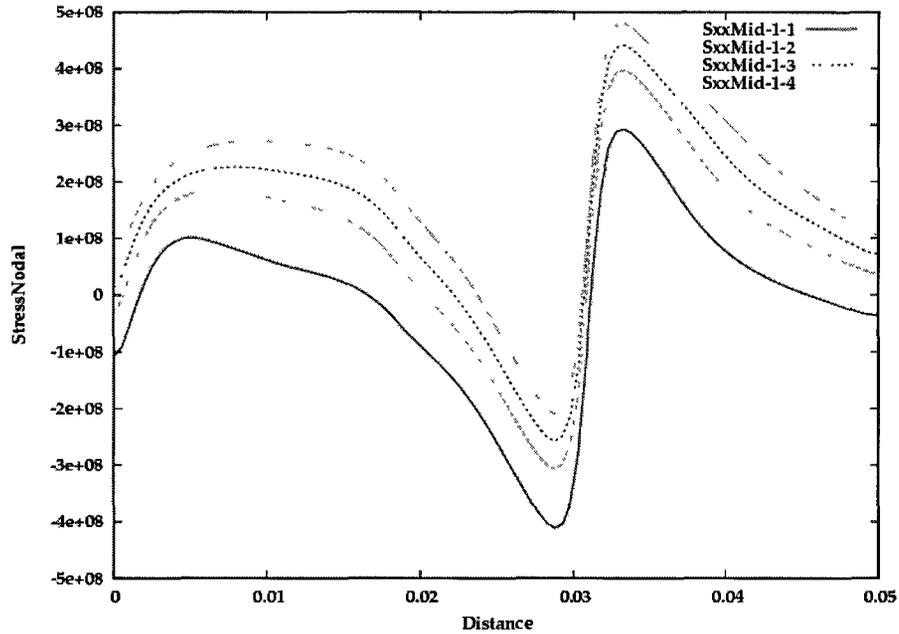


Figure 8.23: Variation in transverse component of stress, σ_{xx} , on the weld path for different tractions when speed-power-double-ellipsoid setting is row 1 in Table 8.3.

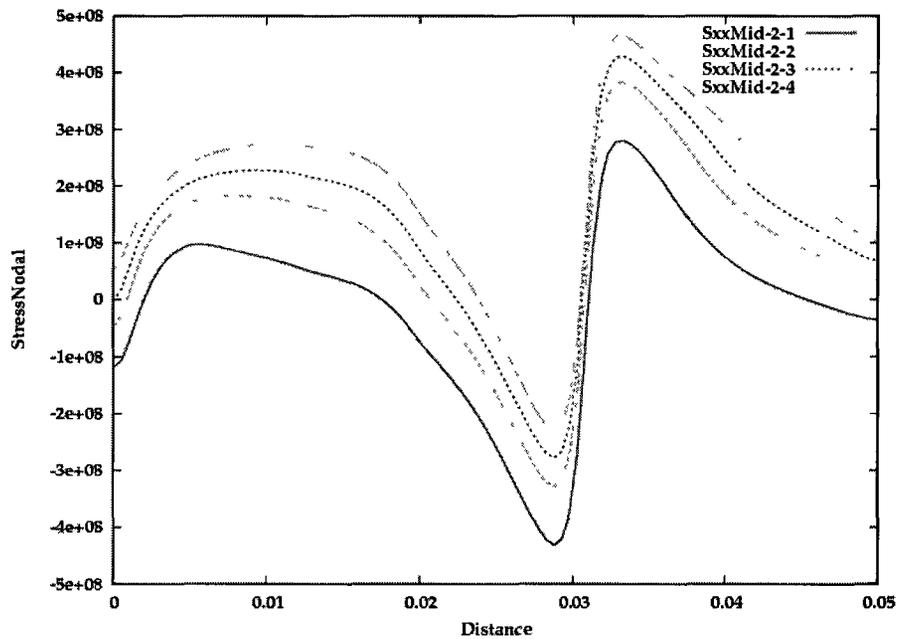


Figure 8.24: Variation in transverse component of stress, σ_{xx} , on the weld path for different tractions when speed-power-double-ellipsoid setting is row 2 in Table 8.3.

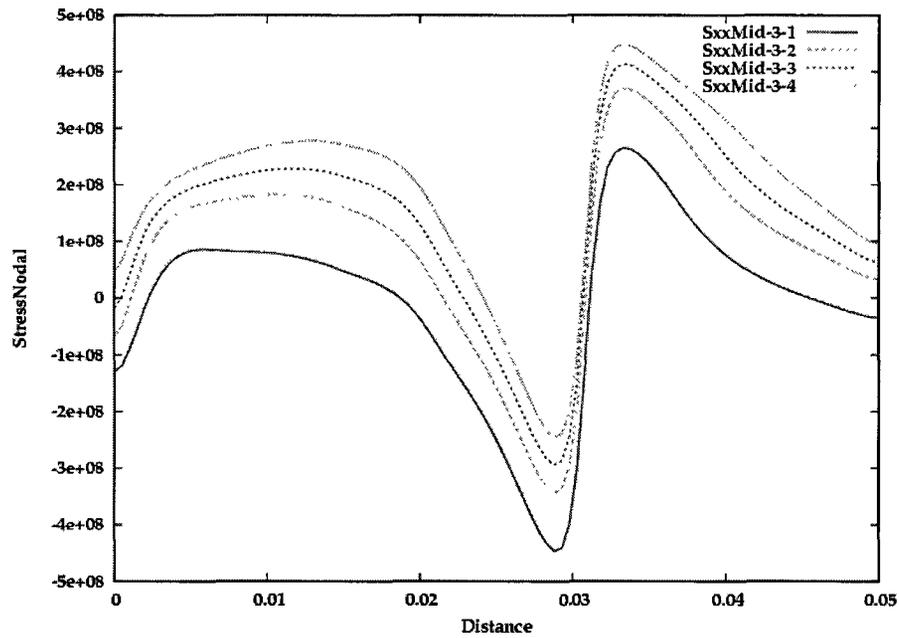


Figure 8.25: Variation in transverse component of stress, σ_{xx} , on the weld path for different tractions when speed-power-double-ellipsoid setting is row 3 in Table 8.3.

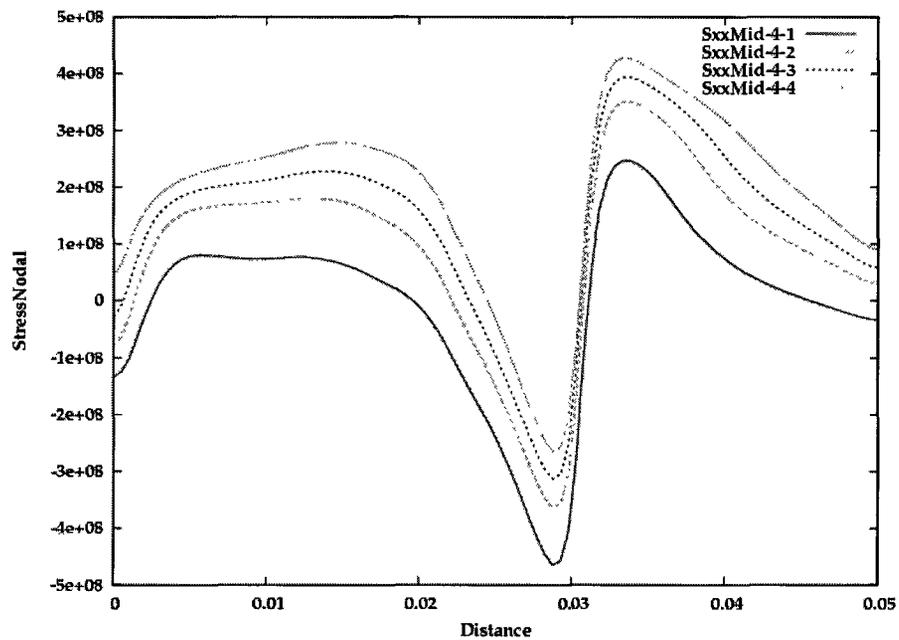


Figure 8.26: Variation in transverse component of stress, σ_{xx} , on the weld path for different tractions when speed-power-double-ellipsoid setting is row 4 in Table 8.3.

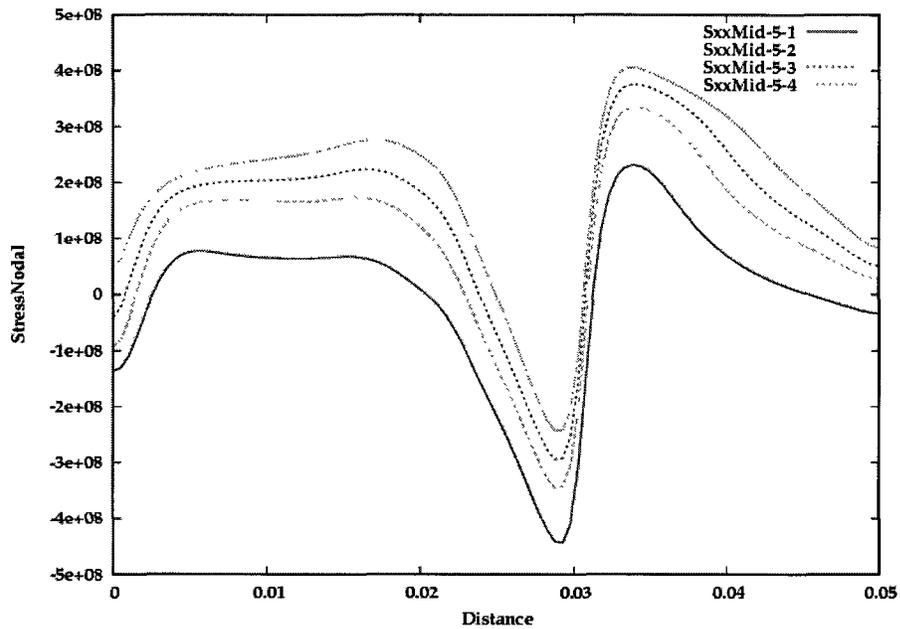


Figure 8.27: Variation in transverse component of stress, σ_{xx} , on the weld path for different tractions when speed-power-double-ellipsoid setting is row 5 in Table 8.3.

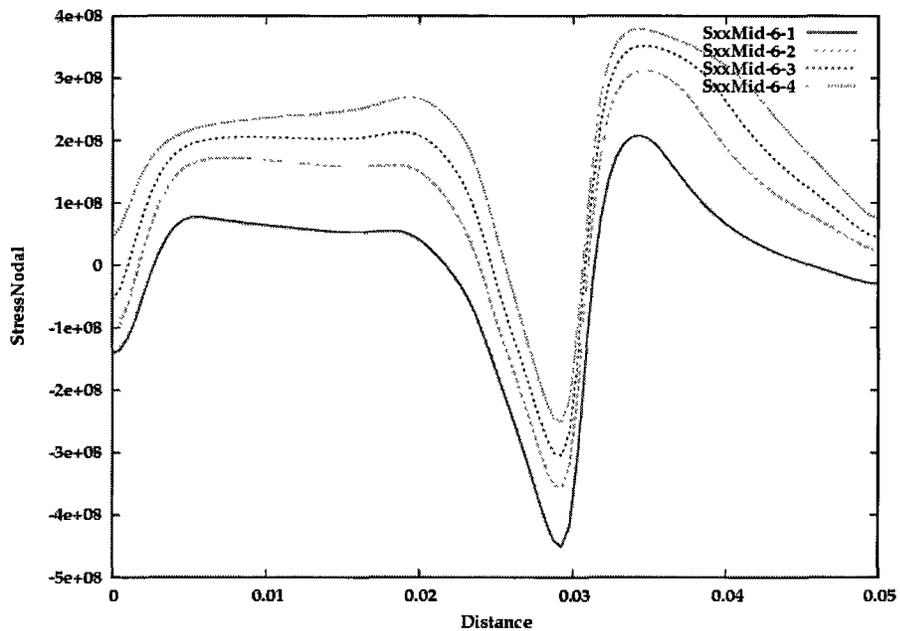


Figure 8.28: Variation in transverse component of stress, σ_{xx} , on the weld path for different tractions when speed-power-double-ellipsoid setting is row 6 in Table 8.3.

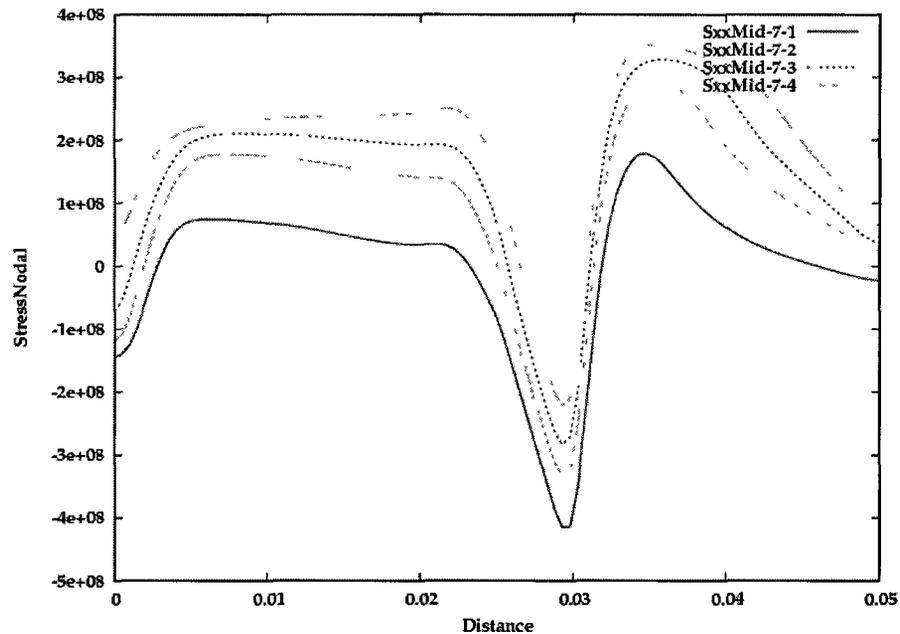


Figure 8.29: Variation in transverse component of stress, σ_{xx} , on the weld path for different tractions when speed-power-double-ellipsoid setting is row 7 in Table 8.3.

the welding speed (with a constant power per unit length) reduces the length of tensile region and therefore reduces the size of the hot cracking region. Fig. 8.23 to 8.29 show that for a given welding speed (with a constant power per unit length), the higher transverse tensile traction increases the length of the tensile region and therefore expands the size of the hot cracking region. The length of the tensile region alone cannot characterize the hot cracking susceptible region because it has no information about the BTR.

A stress vs. temperature plot for nodes on the weld path and behind the weld, can be used to show the transverse tensile stress that lies in the BTR. Although it loses the information about the size and location of hot crack susceptible region, it monitors a criteria to compare the probability of hot cracking in the susceptible region. Fig. 8.10(b) gives the thermal data, and Fig. 8.19 to 8.29 give the stress data for different settings. A post-processor is used to generate the stress vs. temperature plots. Fig. 8.30 compares the stress vs. temperature curves for the best setting, i.e.,

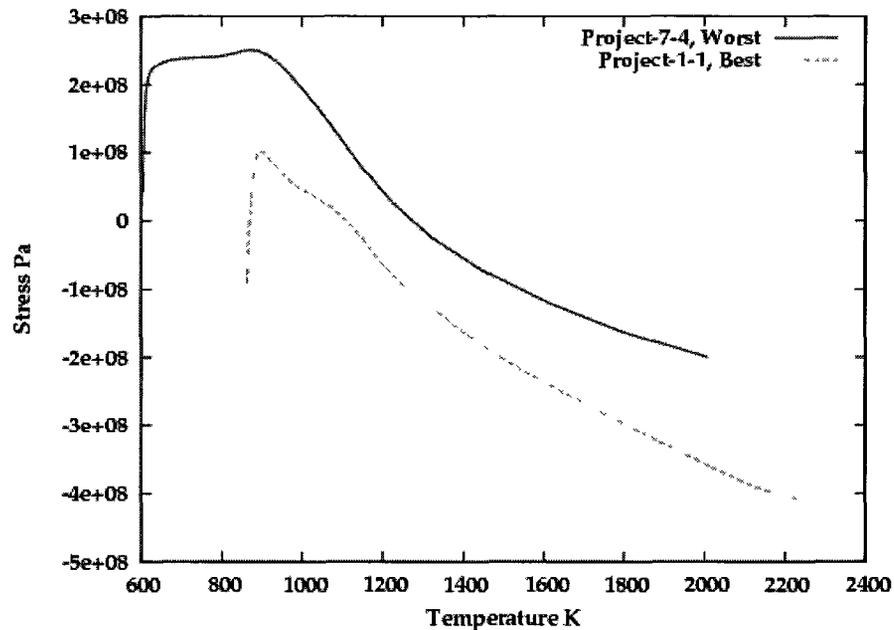


Figure 8.30: Comparison of the hot crack susceptibility between the best setting, i.e., the fastest welding with zero traction, and the worst setting, i.e., the slowest welding with highest traction. The stress component is σ_{xx} .

the fastest welding speed with zero traction, and the worst setting, i.e., the slowest welding speed with highest traction. Fig. 8.31 and 8.32 show the effect of welding speed and traction on the length of tensile stress region behind the weld on the weld path.

For discussion purposes, assuming a BTR range [1000, 1700], Fig 8.33 to 8.38 show the 3Dvisualization of the Gauss points that satisfy the hot cracking conditions for the project 1-1, 1-4, 4-1, 4-4, 7-1, and 7-4 respectively. The radius of the spheres assigned to each Gauss point, shows the magnitude of the stress. The triad shows the weld pool center, the positive z direction is shown along the weld path, and weld moves in negative z direction. For such a selected susceptible region, the possibility of hot cracking differs in each Gauss point because each Gauss point experiences a different level of stress and temperature. For simplicity, assume that the probability of hot cracking for each Gauss point is not sensitive to the nodal temperature value

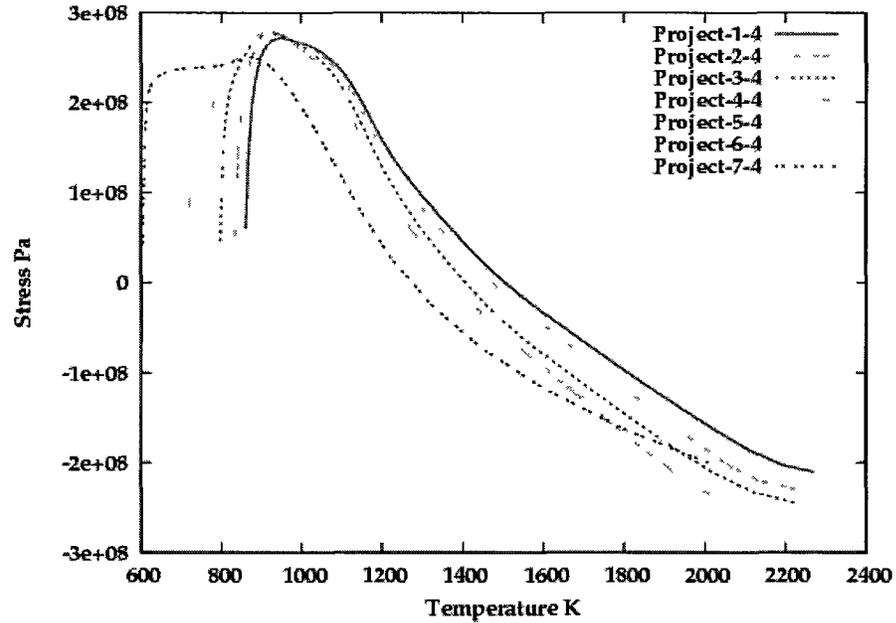


Figure 8.31: Comparison of the hot crack susceptibility for 7 welding speeds when traction is 200 MPa. The stress component is σ_{xx} .

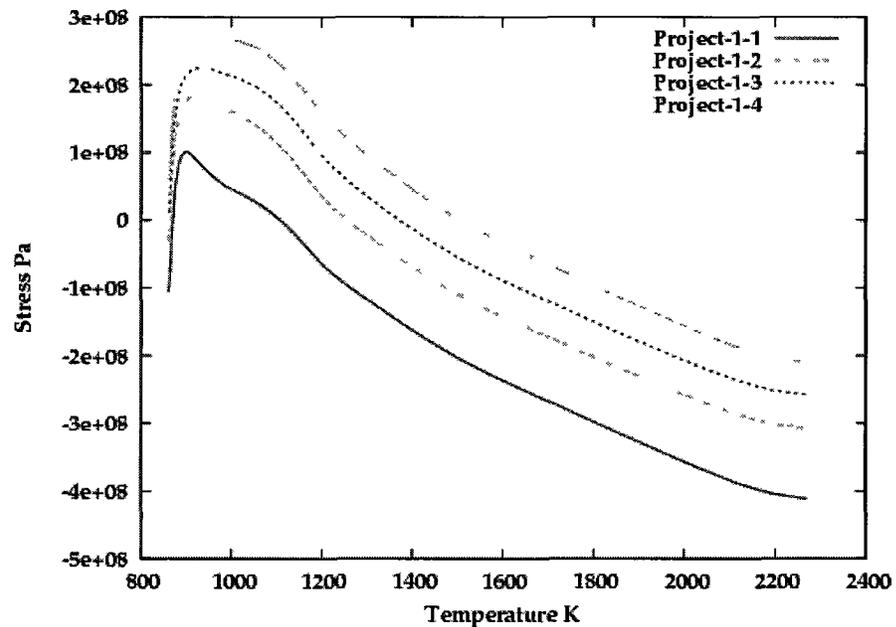


Figure 8.32: Comparison of the hot crack susceptibility for 4 tractions when the welding speed is the fastest. The stress component is σ_{xx} .

when the temperature is in the BTR. In other words, as long as the temperature is in BTR, the stress level on each Gauss point is the dominant factor to define the probability of the hot cracking. Therefore the radius of the spheres in the Fig 8.33 to 8.38 can compare the probability of hot cracking for each Gauss point. In Fig 8.33 the susceptible region is large but there is a low probability of hot cracking for the Gauss points along the weld path. This is shown on the stress vs. temperature plot in Fig. 8.30 that project 1-1 starts tension loading at temperature about 1100 K, i.e., inside the BTR, and the level of tension increases to about 50 MPa at 1000 K, i.e., lower bound of the BTR. However, in 3D visualization (Fig 8.33) there exist few big spheres on both sides of weld path in the susceptible region that show the high probability of hot cracking not on the weld path but on the sides of the weld path. In Fig 8.38 the susceptible region is small but there is a higher probability of hot cracking for the Gauss points in this region in comparison with Fig 8.33. This also is shown on the stress vs. temperature plot in Fig. 8.30 that project 7-4 experiences higher tension at each temperature in compare to project 1-1. Comparing the 3D visualization of Fig 8.33 and Fig 8.38, susceptible region is smaller in project 7-4 but the probability of hot cracking is higher in this small region than project 1-1. The susceptible region is larger in project 1-1 but most of the Gauss points are not very likely to crack. There exist few Gauss points in project 1-1 on the sides of the weld path that might crack with higher probability of cracking than the Gauss points in project 7-4. The 3D results show the almost empty susceptible region for project 7-1. In the author's opinion, a post-processor that provides such 3D transient field, is a reliable way to judge about the susceptibility of hot cracking from a computational analysis.

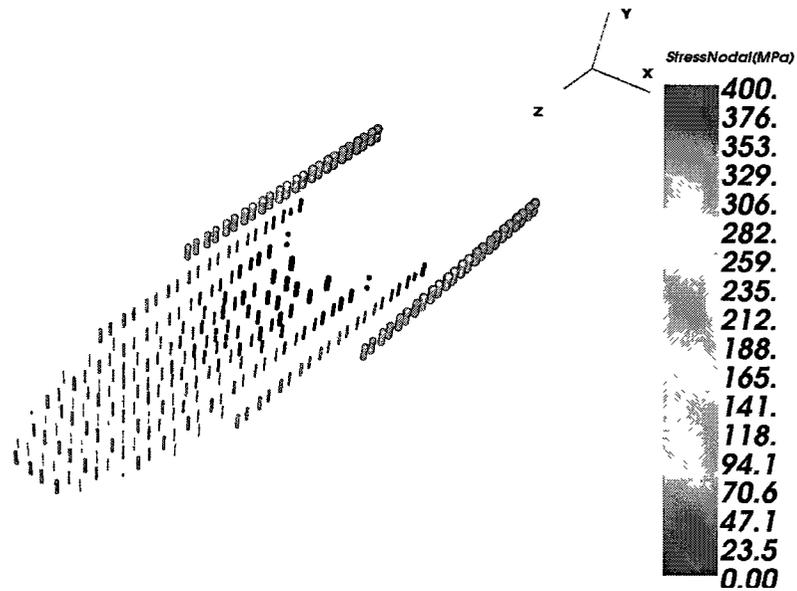


Figure 8.33: 3D visualization of the Gauss points that satisfy the hot cracking conditions for BTR [1000, 1700] for the project 1-1.

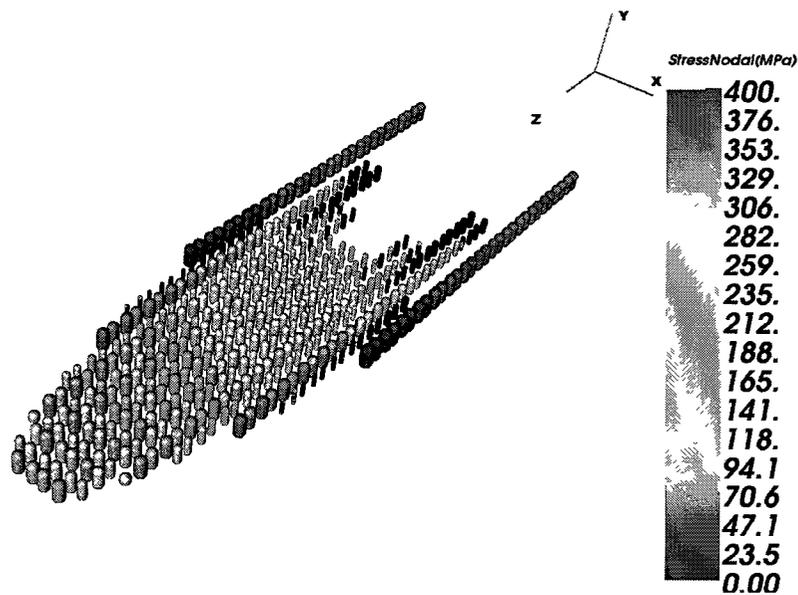


Figure 8.34: 3D visualization of the Gauss points that satisfy the hot cracking conditions for BTR [1000, 1700] for the project 1-4.

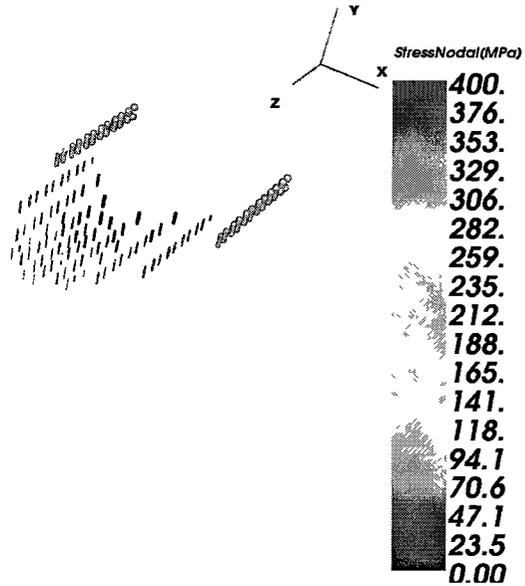


Figure 8.35: 3D visualization of the Gauss points that satisfy the hot cracking conditions for BTR [1000, 1700] for the project 4-1.

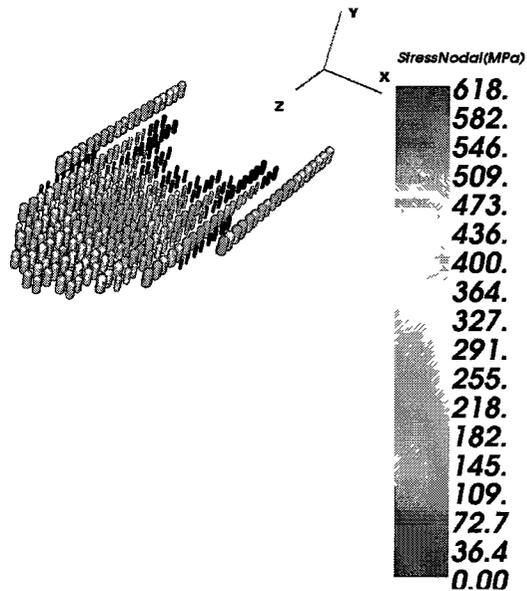


Figure 8.36: 3D visualization of the Gauss points that satisfy the hot cracking conditions for BTR [1000, 1700] for the project 4-4.

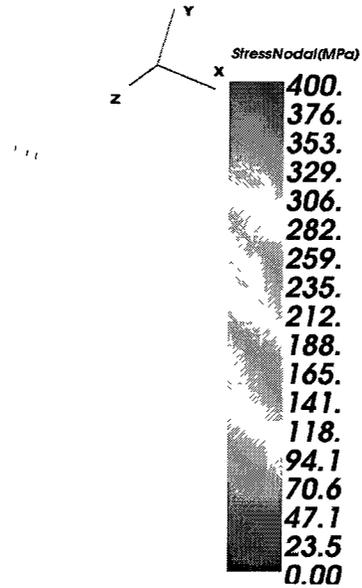


Figure 8.37: 3D visualization of the Gauss points that satisfy the hot cracking conditions for BTR [1000, 1700] for the project 7-1.

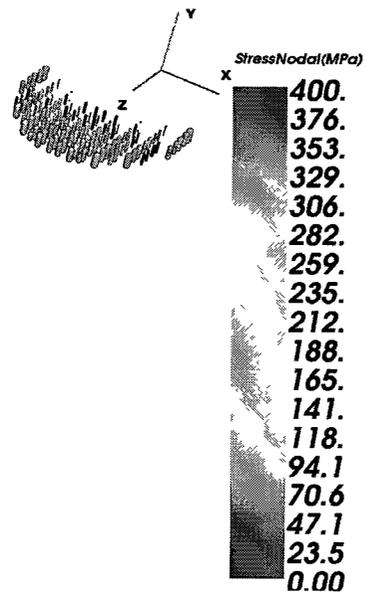


Figure 8.38: 3D visualization of the Gauss points that satisfy the hot cracking conditions for BTR [1000, 1700] for the project 7-4.

Assuming the BTR for 316 from [44], i.e., [1200, 1400] K, the slow weld does hot crack and the fast weld does not hot crack. Post processing thermal and stress data over all Gauss points generates a 3D field of hot crack susceptibility for every time step. The post-processor checks three conditions for hot cracking; 1) temperature in cooling thermal profile 2) temperature in the BTR, [1200, 1400] °K in this case, and 3) tensile stress envelope, over all Gauss points. The fourth condition which is the effect of strain rate, is assumed CST equal to zero. Fig. 8.39 and 8.40 show 3D visualization of the Gauss points that satisfy the hot cracking conditions for BTR \in [1200, 1400] for the worst case, i.e., slowest weld with highest traction. The corresponding field is empty for the best setting, i.e., fastest welding with zero traction.

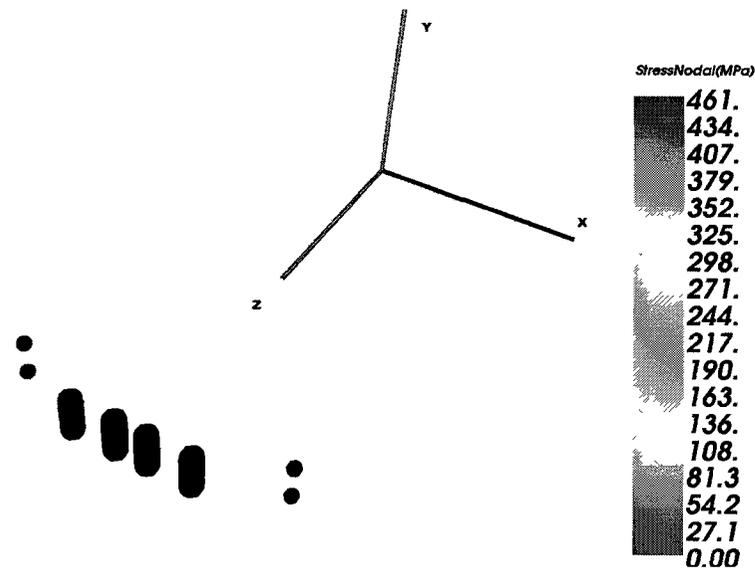


Figure 8.39: 3D visualization of the Gauss points that satisfy the hot cracking conditions for the worst setting, slowest weld with highest traction. The radius of the spheres shows the transverse stress magnitude.

8.8 Closure of the Analysis

Given a CWM framework that can automate the implementation of a great number of analyses, then control, optimization, and sensitivity analyses can be implemented

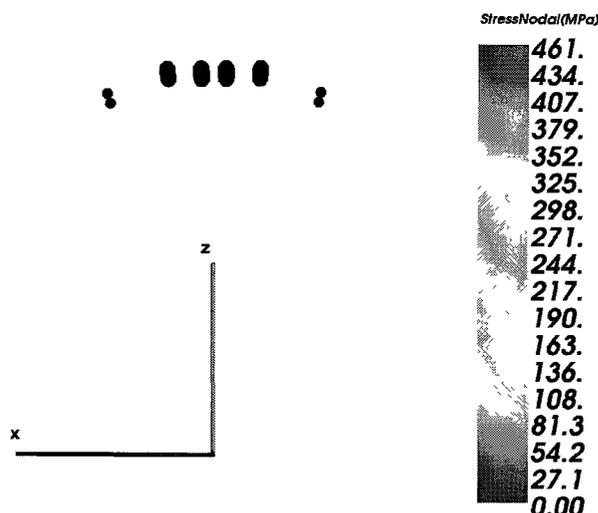


Figure 8.40: Top view of a 3D visualization of the Gauss points that satisfy the hot cracking conditions for the worst setting, slowest weld with highest traction. The radius of the spheres shows the transverse stress magnitude.

for CWM problems. Post-processor algorithms that use the transient 3D thermal and stress data to identify the 3D transient hot crack susceptible region in a welded structure.

The sensitivity analysis of hot crack susceptibility on the specimen taken from [44], required 9 DOE-matrices with 13 analyses each to control the 4 double ellipsoid shape parameters for different speeds, and required 4 DOE-matrices with 7 analyses each for the sensitivity analyses wrt traction and welding speed ¹.

A hot cracking algorithm used the 3D transient thermal and stress data over all Gauss points for crack susceptibility. For each Gauss point in the domain the hot cracking algorithm checks if the temperature is in the cooling scheme, the temperature lies in the BTR range, and the transverse stress is tensile. Such post processor is a reliable way to judge about the susceptibility of hot cracking from a computational analysis.

¹The base project comparable to the test done in [44] is carried out by other members of the research group. The base project is employed to implement the control and sensitivity analyses with time spent only for DOE matrix design and almost no time for setup.

For BTR in $[1200, 1400] \text{ } ^\circ K$, the sensitivity analysis shows that increasing the tensile traction transverse to the weld path and reducing the weld speed increase the chance of hot cracking. A tensile-stress vs. temperature along the weld path and behind the weld is used to monitor the hot crack susceptibility. Comparing the results indicates that increasing the welding speed at constant power per unit length can reduce the risk of hot cracking.

Chapter 9

A Discontinuous Combinatorial Design Space Exploration; Surrogate Modelling

Project: Combinatorial Optimization of Weld Sequence by Using a Surrogate Model to Mitigate a Weld Distortion

Choosing an optimal sequence from the set of all possible combinations of a weld's sub-passes has been always a challenge for designers. The solution of such combinatorial optimization problems is limited by the available resources. For example, having n sub-passes leads to choosing from $2^n \times n!$ possible combinations of the sub-passes, e.g., 46,080 for $n=6$. Although the prior knowledge can reduce this space, still it is not practically feasible to choose the optimal sequence by evaluating a set of possible combinations either experimentally or by simulation models. Using a surrogate model based on a simulation model is to find the solution in the space of all possible combinations with a significant decrease in computational expenses. The surrogate model used in this analysis is constructed for a discontinuous design space. This is quite different concept from surrogate functions for continuous functions which is constructed from interpolation functions, regression approximation or gradients. The discontinuous surrogate model constructs an approximation model from some combinations of

sub-sequences of a more expensive model to mimic the behaviour of the expensive model as closely as possible but at a much lower computational cost. This surrogate model could be used to approximate the behaviour of the other weld sequences. The optimal solution or solutions determined by the surrogate model can be tested by evaluation of the expensive model. In this analysis, a surrogate model is demonstrated that minimizes the distortion in a pipe girth weld with six sub-passes by analyzing only 14 combinations of sub-passes from total of 48 possible combinations. A comparison between the results of the surrogate model and the full transient FEM analysis of all possible combinations shows the accuracy of the surrogate algorithm/model.

9.1 Analysis Background

Distortion in welded structures results from the plastic strain that is driven by the expansion and contraction of the weld metal and adjacent base metal during the heating and cooling cycle of the welding process. During this heating and cooling cycle, many factors affect shrinkage of the metal and lead to distortion, such as physical and mechanical properties that change as heat is applied [31]. Very good simulation codes are now available to predict the behaviour of a welded structure based on 3D transient temperature and thermal stress-strain analysis. However simulation codes are limited by available resources and using them for problems involving a great number of projects such as combinatorial optimization problems might not be feasible.

There are efforts to develop more affordable approaches to have sufficient and reliable level of understanding of the behaviour of structures while welding.

One approach is to use a fast simulation code that captures the most dominant physics of the problem. Although such a code or algorithm loses accuracy, the hope is that it provides a useful approximation of the behaviour. In many design cases, the designers can make a decision based on this rough approximation of the behaviour.

Good examples of these algorithms are explained in [49] and [50] for a ‘Fast Weld’ and ‘Super Fast Weld’ algorithm that analyzes the weld displacement based on elastic-thermal-strain only. They show a good estimate of a full 3D transient analysis for a truck’s axle bridge with 4 welds but are as much as 30 time faster than a normal analysis.

Another approach is to construct an approximate model from the behavior computed by simulation codes and use this approximate model to predict the behaviour for all the other cases. Regression models are one such an approximate model. Reference [29] constructs a regression model for the residual stress approximation in a bead-on-plate weld. An analysis with a sample space size of 10,000 points is done very quickly using this regression model.

Surrogate models, called also meta models or emulators, can be another alternative for the approximate model. These models mimic the behaviour of the simulation model as closely as possible while being computationally cheaper to evaluate [51]. A surrogate model is employed by Voutchkov et al. in [7] to optimize the weld sequence for a tail bearing housing that mounts the engine to the body of an aircraft. Weld sequence optimization is an instance of a discontinuous design space in which all solutions are isolated points. There is no notion of continuity, distance or topology between solution points, i.e., weld sequences. Therefore the surrogate function used in [7] and in this analysis to optimize weld sequences for the pipe’s girth weld, is quite different from the surrogate function for continuous functions that is described in [52] which is constructed from interpolation functions, regression approximation or gradients to be used for continuous function spaces. In case of Voutchkov et al. [7], there are 6 sub-passes and each sub-pass can be done in two directions. Voutchkov et al. must choose the weld sequence from 46,080 possible weld sequences to minimize the distortion. Voutchkov et al. picked 27 sequences from the combinatoric space of 46,080 combinations to analyze with an expensive model that required 35 hours of

CPU for each analysis. Then they used those 27 solutions to construct a surrogate model that required only 14 flops for each sequence, i.e., 7 real scalar multiplies and 7 real scalar additions for each sequence. This requires less than 1 minute CPU time for each sequence. They solved the remaining 46,053 weld sequences with this surrogate model. They show that the solution of the surrogate model was very accurate by comparing the surrogate solution with the FEM solution.

9.2 Girth Weld in Pipe

Pipe girth welds are widely used in a variety of engineering applications such as oil and gas industries, nuclear and thermal power plants and chemical plants.

In this analysis, the girth weld connects two pipes of length 356 [mm], wall thickness 17.5 [mm] and outer diameter 324 [mm]. The pipes are part of a big pipe loop and therefore constraints are applied in the analysis to approximate the effect of the rest of the loop. The cross section and mesh are shown in Fig. 9.1. The element type is an 8-node brick. The number of volume elements and volume nodes are 24,064 and 36,228 respectively. The original setting has 5 weld passes or layers of weld to fill the joint between the connected pipes. The analysis of the first and second layer is used to construct the surrogate model. This is because layers 2 to 5 are very similar.

The weld procedure for each layer is given in Table 9.1. Each layer is divided into three sub-passes and each sub-pass covers 1/3 of the whole layer. The two pipes are tack-welded prior to welding at three points and each sub-pass starts from one tack-weld and ends on the next one. Here, the letters a, b, and c denote the sub-passes from tack-weld 1 to tack-weld 2, tack-weld 2 to tack-weld 3 and tack-weld 3 to tack-weld 1 respectively. Each sub-pass can be done either clockwise, here a negative sign denotes the counterclockwise direction. Therefore there is a sequence of 3 sub-passes chosen from a, -a, b, -b, c, and -c for each layer. This is shown in

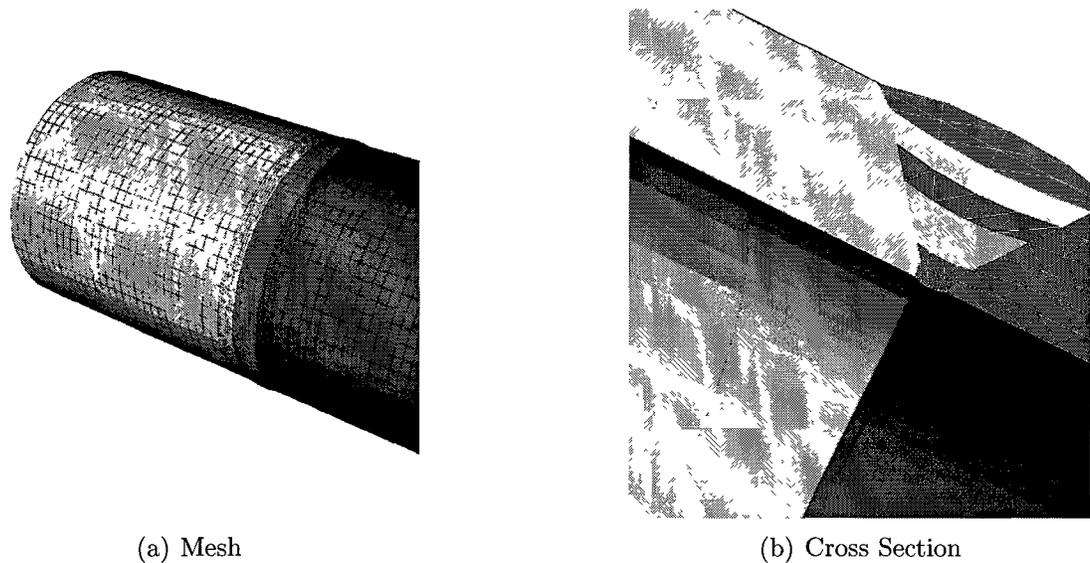


Figure 9.1: The cross section and mesh employed in this analysis. Wall thickness is 17.5 [mm].

Fig. 9.2. There are delay-times of 2 minutes between each sub-pass and the pipes are cooled down to the ambient temperature after welding is completed. The timing used in the simulation was 114 seconds welding in addition to 120 seconds cooling for each sub-pass of the first layer and 128 seconds welding in addition to 120 seconds cooling for each sub-pass of the second layer. There is a long cooling time at the end of the third sub-pass of the second layer to fully cool down to the ambient temperature so that the total time including welding and cool-down is 5,700 seconds.

Table 9.1: Weld procedure used for each layer of weld in this analysis.

Parameter	Layer 1	Layer 2
Welding Speed [mm/s]	2.66	2.38
Voltage [v]	29	17
Current [amp]	165	155
Power Efficiency	0.9	0.70



Figure 9.2: The notation of sub-passes for the first and second layer. They can have negative sign to show the change in direction.

9.3 Surrogate Model

Optimization requires evaluation of a design objective and the constraints as a function of design variables. One approach is to construct an approximate model such as a surrogate model that mimics the behaviour of the simulation model as closely as possible while being computationally less expensive to evaluate [51].

Welsh et al. [53] show an approach based on fitting response surfaces to data collected by evaluating the objective and constraint functions at a few points. They observed how the function typically behaves such as how much the function tends to change as each parameter changes. In the prediction for a new point, they are essentially computing the function value that is most consistent with this estimated typical behaviour. Voutchkov et al. [7] developed a surrogate model based on this idea for weld sequence optimization to make it affordable by selecting 27 sequences in the space of 46,080 possible combinatorial weld-sequence space.

A surrogate-based optimization typically includes the following steps;

1. Design of experiments (DOE) matrix.
2. Evaluate the DOE matrix with an accurate model.
3. Use this evaluation to train a surrogate model to approximate the desired relationships.

4. Evaluate all combinations using the trained surrogate model.
5. Select the optimum result from the surrogate and evaluate it using the accurate model.
6. If the result is not good enough, add it to the existing DOE matrix results and re-train the surrogate model.

For the case of combinatorial optimization in [7], the total permanent displacement at one distinguished point from welding is described as a sum of the nodal displacements from each individual weld after the cool down. Dependency on weld position and direction is added by constructing a mathematical series as in Eq. 9.1 that captures increments due to the position and direction of each weld. For the first weld of the sequence, the displacement caused is independent of subsequent welds and they term it a *main effect*.

$$D = \sum_{i=1}^n (M_{w_i} + \Delta(w, i)) \quad (9.1)$$

D : Total displacement.

M_{w_i} : Main effect of weld w_i

$\Delta(w, i)$: Correction due to the position and direction of weld w_i in a sequence.

To accommodate the correction factor $\Delta(w, i)$, they start picking the combination from the space of all possible combinations such that each pair in a sequence occurs at least once regardless where the pair occurs, i.e., called R_2'' -matrix in [7]. Voutchkov et al. picked 27 sequences in this fashion. Then they ran an FEM for analyses for each of the 27 sequences.

Two definitions are taken from [7] for the terms Order and Type. Order denotes the size of the substring and is designated by subscripts. For example, substring (a

-b, c) has the order of 3 and is denoted by $R_3(a -b, c)$. Type denotes whether the sequence is available in the sample space (type 1) or if it is mimicked from some other sequence in the sample space (type 2). Type 1 is denoted by R' and type 2 by R'' . To construct the surrogate model, they used the following priority list for training the surrogate model [7]:

1. Start with the full string (highest order)
2. Search for any occurrences of type 1 of this order.
3. If nothing found, search for any occurrences of type 2 of this order.
4. If nothing found, truncate the string from left and update the order
5. Repeat until the main effect is included.

This priority list was tried for training the current model in case of the girth weld and the approximation was found to be weak compared to the exact results. Fig. 9.3(a) shows a comparison between the FEM and surrogate approximation based on the priority list algorithm in [7]. Their recipe is based on the assumption that most of the effect comes from the first sub-pass in the sequence. This assumption works well for the case in [7], a tail bearing housing, because the maximum distortion/displacement occurs at a place far from each sub-pass with no significant change in this distance from sub-pass to sub-pass. In the case of a pipe girth weld, the maximum distortion/displacement occurs at the weld or around it and therefore the distance has a very significant effect. That makes the recipe used in [7] inaccurate for distortion in a pipe girth weld.

Therefore another algorithm was developed to make a priority list based on the physics of the girth weld to train the surrogate model. Fig. 9.3(b) shows a comparison between the FEM and the surrogate approximation based on the priority list employed in this analysis.

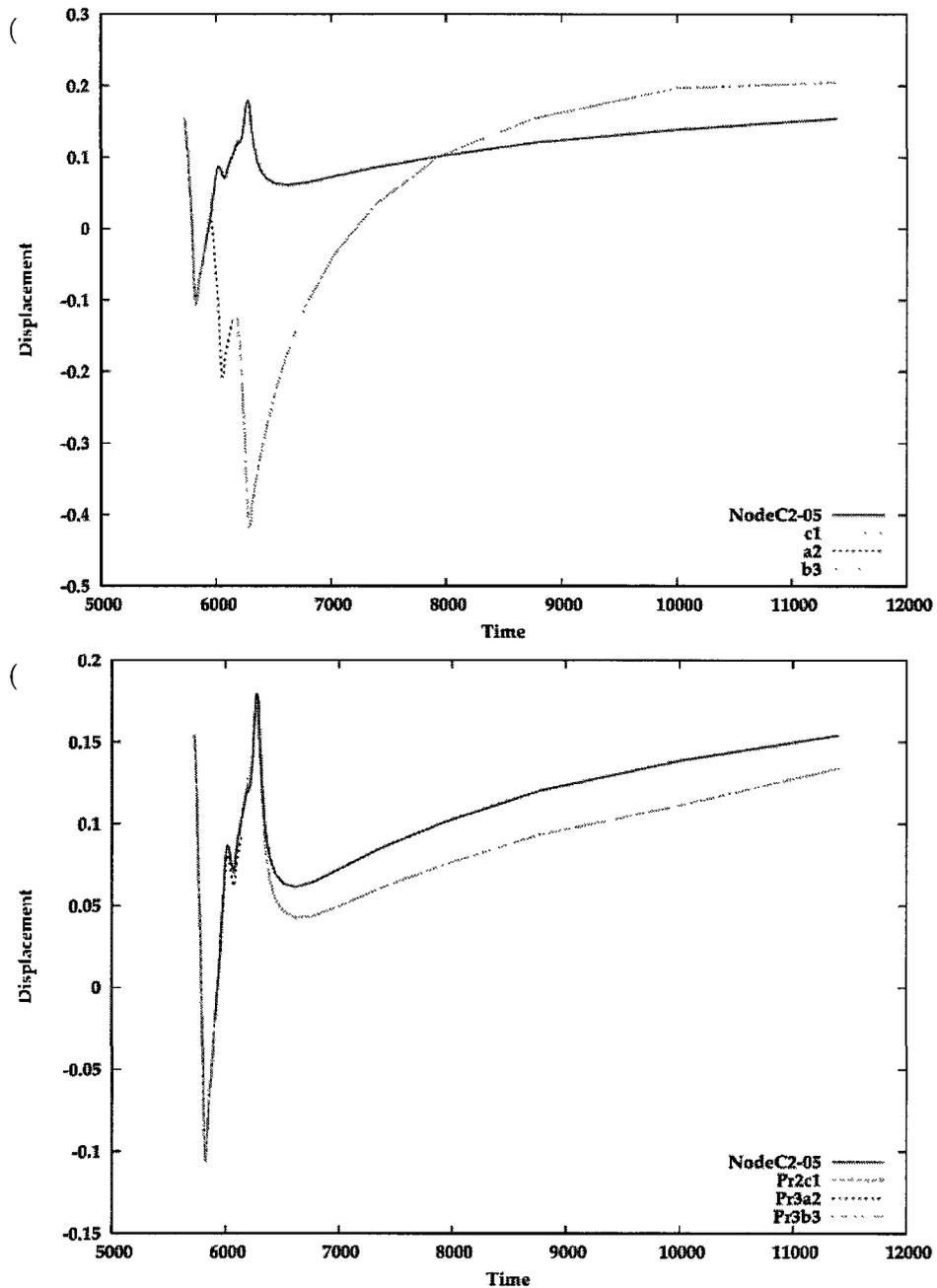


Figure 9.3: Comparison between the surrogate approximation constructed based on the algorithm presented in [7], 9.3(a), and the surrogate approximation constructed based on the algorithm developed in this analysis, 9.3(b). The red line shows the exact result from FEM and the other colors show each term in the surrogate approximation.

9.4 Computational Setups

The full computational model that includes transient thermal and stress analysis is analyzed with VrWeld software [21].

9.4.1 Thermal simulation

The power density distribution function Q [w/m^3] ‘Added Filler Metal’ heat source model [23] was used and the heat source parameters; T_{part} , $T_{previous}$ and $T_{interior}$ are set to $1700\text{ }^\circ\text{K}$, $1800\text{ }^\circ\text{K}$ and $1900\text{ }^\circ\text{K}$. A convection boundary condition generated a boundary flux q [w/m^2] on all external surfaces. This analysis uses the convection coefficient $h_c = 10$ [$\text{w}/\text{m}^2\text{K}$] and ambient temperature of $T_{ambient} = 300$ [$^\circ\text{K}$]

The time step length while welding was chosen so that one time step was required to travel one element along the weld path. Filler metal was added as the welding arc moved along the weld path, i.e., the FEM domain changed in each time step during welding. After each weld pass was completed, the time step length was exponentially increased by a factor of 1.2 per time step for the cooling cycle until the next weld pass started.

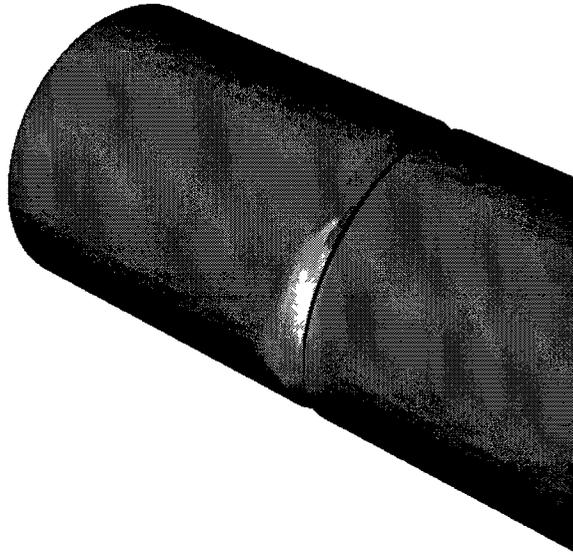
The CPU time was 445 seconds and 613 seconds for the thermal analysis of the first layer and second layer respectively.

Fig. 9.4 illustrates a snapshot of the transient temperature field when the weld is in the middle of the first sub-pass of the first layer.

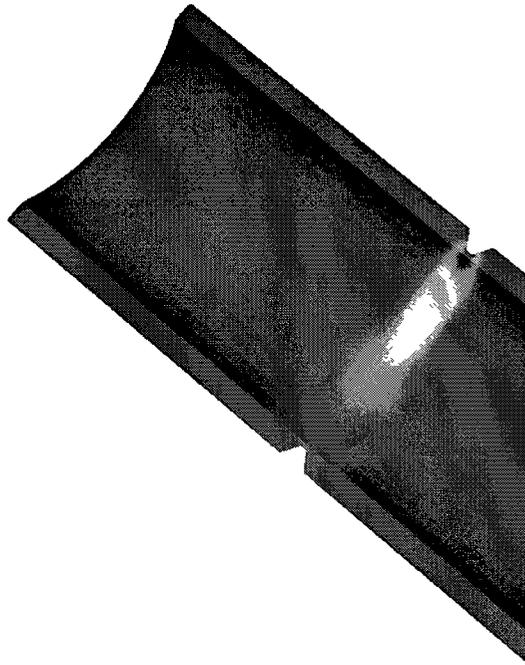
Fig. 9.5 illustrates a snapshot of the transient temperature field when the weld is in the middle of the first sub-pass of the second layer.

9.4.2 Stress simulation

The system is solved using a time marching scheme with time step lengths of approximately 3 seconds during welding and an exponentially increasing time step length

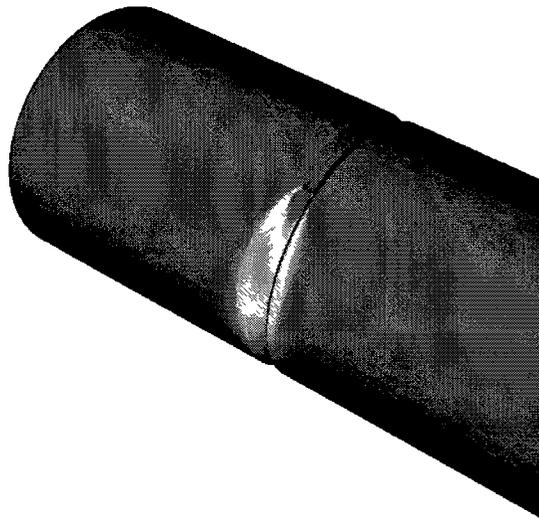


(a) First layer, First Sub-pass

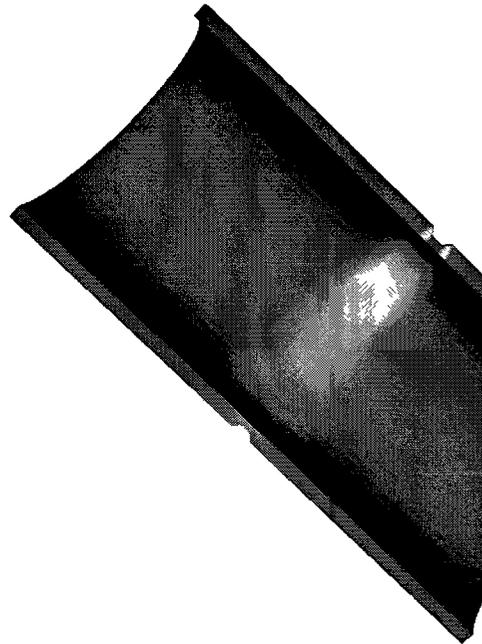


(b) First layer, First Sub-pass, Cross Section

Figure 9.4: A snapshot of the thermal result that shows the transient temperature field when the weld is in the middle of the first sub-pass of the first layer. Wall thickness is 17.5 [mm].



(a) Second layer, First Sub-pass



(b) Second layer, First Sub-pass, Cross Section

Figure 9.5: A snapshot of the thermal result that shows the transient temperature field when the weld is in the middle of the first sub-pass of the second layer. Wall thickness is 17.5 [mm].

when welding stops.

The total CPU time was 15 hours and 20 hours for the thermal analysis and stress analysis of the first layer and the second layer of the girth weld respectively.

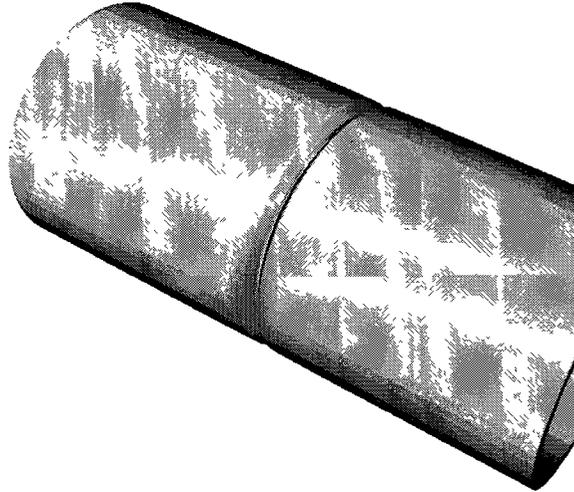
Fig. 9.6 illustrates the displacement and axial stress field when the first layer of weld finished using the sequence (a, b, c) and cooled down to the ambient temperature.

Fig. 9.7 illustrates the displacement and axial stress field when the second layer of weld finished using the sequence (a, b, c) and cooled down to the ambient temperature. This layer follows the weld shown in Fig. 9.6.

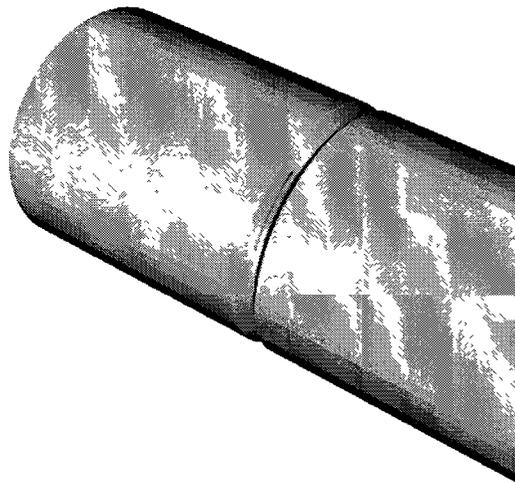
9.5 Surrogate Algorithm for Girth Weld

This section describes the surrogate algorithm constructed and used for this pipe girth weld. In the first layer, there is an inherent symmetry in the pipe due to the circular shape so that the space of possible combinations of sequence contracts to a significantly smaller space. For example, the sequences (a, b, c), (b, c, a) and (c, a, b) are the same for the first layer if there is no initial state. In fact the sequence (b, c, a) and (c, a, b) are 120° and 240° rotation of the sequence (a, b, c) respectively with no effect on the physics of the problem unless there is an initial state that removes the symmetry assumption. Therefore there is no need to permute the first sub-pass and it can be either a, b, or c. The sub-pass a was selected for the first sub-pass and permute the other sub-passes: b, -b, c and -c. Table 9.2 shows all possible sequences for the first layer of weld regarding the symmetry in pipe's shape.

Having the total combination of size 6 for the first layer changes the strategy for solving the problem considering the trade-off between the accuracy and available resources.

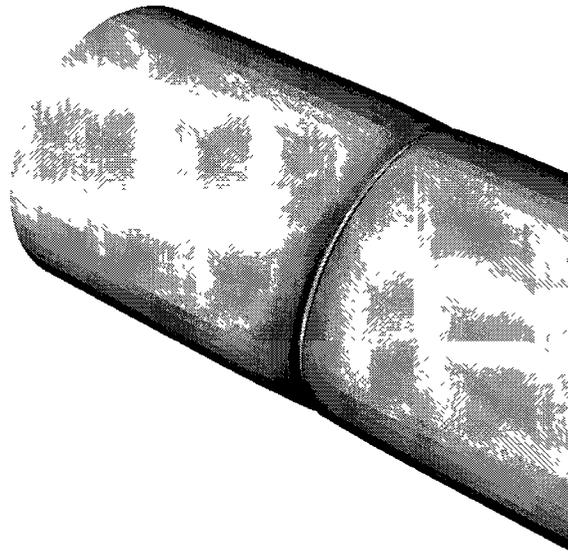


(a) First layer, Displacement, Visualization x10

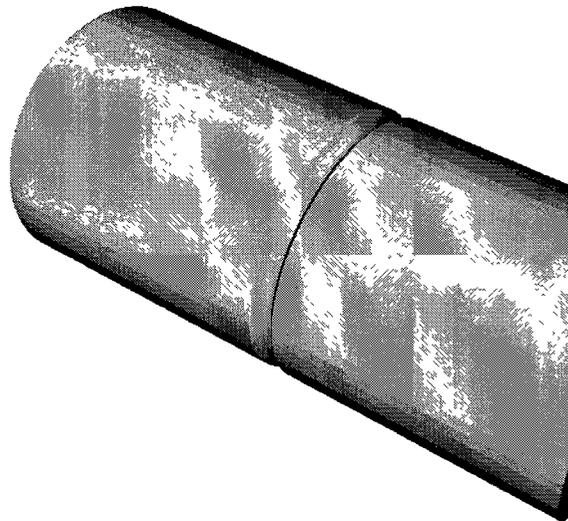


(b) First layer, Axial Stress

Figure 9.6: Total displacement and axial stress field when the first layer of weld finished using the sequence (a, b, c) and cooled down to the ambient temperature. Wall thickness is 17.5 [mm].



(a) Second layer, Displacement, Visualization x10



(b) Second layer, Axial Stress

Figure 9.7: Total displacement and axial stress field when the first layer of weld finished using the sequence (a, b, c) and cooled down to the ambient temperature. This layer follows the weld shown in Fig. 9.6. Wall thickness is 17.5 [mm].

Table 9.2: All possible sequences for the first layer.

Number	First Sub-pass	Second Sub-pass	Third Sub-pass
1	a	b	c
2	a	b	-c
3	a	-b	c
4	a	-b	-c
5	a	-c	b
6	a	-c	-b

This suggests using the full transient FEM analysis in order to find the global-minimum distortion in the total space of combinations. Each analysis takes 15 hours and employing the multi-project mode in VrWeld using 3 CPUs to run all settings defined in Table 9.2 in parallel. The analyses were completed in 30 hours and the maximum displacements are reported in Table 9.3.

Table 9.3: Maximum displacement for the first layer [mm].

Run No.	During the Process	End of the Process
1	0.447	0.412
2	0.431	0.418
3	0.485	0.420
4	0.4767	0.391
5	0.447	0.429
6	0.488	0.4882

Designers try to minimize the maximum displacement at the end of process and therefore run number 4 was picked from Table 9.3. This results from the sequence

(a, -b, -c). All stress states including the nodal stresses and displacements from this sequence were then set as initial state for the analysis of the second layer. This saves time by avoiding the first layer re-analyses when analyzing the second layer.

The symmetry assumption could not be used for the second layer. Therefore all combinations of a, -a, b, -b, c, and -c had to be permuted. There are many algorithms to generate these permutations. In this case, the algorithm used, was:

1. Generate the permutation of (a, b, c) that generates 6 cases.
2. Change all “a” to “-a” that generates another set of 6.
3. Change all “a” and “b” to “-a” and “-b” that generates another set of 6.
4. Change all “a” and “b” and “c” to “-a” and “-b” and “-c” that generates another set of 6.
5. Change all “b” to “-b” that generates another set of 6.
6. Change all “a” and “c” to “-a” and “-c” that generates another set of 6.
7. Change all “c” to “-c” that generates another set of 6.
8. Change all “b” and “c” to “-b” and “-c” that generates another set of 6.

Finally there are 8 sets of 6 for a total of 48 possible combinations as shown in Table 9.4. However it is not necessary to generate all combinations to construct the surrogate model and the algorithm can stop generating combinations as soon as it has all pairs of sequences regardless of the position. For example, in [7], the total combinatoric space has a size of 46,080 sequences and they stop after picking 27 sequences out of 46,080. They introduced the R_2'' -matrix that keeps track of each pair regardless of the position in the sequence. Here, the same approach was used to

control the picking algorithm. The same notations as reference [7] are employed for this analysis.

Having the 48 sequences shown in Table 4, makes the CPU time for the FEM analyses too long. If one more sub-pass is added, the size of the total combinatorial space goes to 384 following the formula $2^n \times n!$. This encourages use of approximate models such as the surrogate one.

To pick the sample space from the total space of combinations, start picking randomly from Table 9.4 one-by-one and fill out the corresponding entries in the R_2'' -matrix. This process continues until all entries have been filled in the R_2'' -matrix. The current sample space has 14 sequences as shown in Table 9.5.

The R_2'' -matrix idea from [7] is simplified as it keeps the occurrence by assigning 1 or 0 to the matrix instead of putting the correction values into the matrix. The reason for using this matrix is to keep track of sequences in each picking step. The value of correction is computed in another part of our algorithm. This is more efficient because all corrections need to be computed. It can be done and cached in temporary memory for the favorite case at any time and eliminating the need to store all the data. The R_2'' -matrix is shown in the Table 9.6 and the number in position $[i, j]$ shows the number of occurrences of sub-pass j after sub-pass i . The diagonal element has no meaning and hence the values are set to zero.

Another table is used to keep track of each sub-pass occurrence in the first, second, and third position called R_1' -matrix as shown in Table 9.7. This matrix should be filled with none-zero values at the end of the picking algorithm. This matrix shows, for instance, sub-pass (a) occurs at the first position 4 times and so on.

Each FEM analysis of one sequence took 20 hours and using parallelization on 4 cores, the results for 14 sequences were computed in 80 hours. The priority list employed to construct the surrogate model is based on the physics of the problem such that the closest sub-pass has the dominant effect on displacement of each node.

Table 9.4: All sequences for the second layer.

No	First Sub-pass	Second Sub-pass	Thrd Sub-pass
1	a	b	c
2	a	c	b
3	b	a	c
4	b	c	a
5	c	a	b
6	c	b	a
7	-a	b	c
8	-a	c	b
9	b	-a	c
10	b	c	-a
11	c	-a	b
12	c	b	-a
13	-a	-b	c
14	-a	c	-b
15	-b	-a	c
16	-b	c	-a
17	c	-a	-b
18	c	-b	-a
19	-a	-b	-c
20	-a	-c	-b
21	-b	-a	-c
22	-b	-c	-a
23	-c	-a	-b
24	-c	-b	-a
25	a	-b	c
26	a	c	-b
27	-b	a	c
28	-b	c	a
29	c	a	-b
30	c	-b	a
31	-a	b	-c
32	-a	-c	b
33	b	-a	-c
34	b	-c	-a
35	-c	-a	b
36	-c	b	-a
37	a	b	-c
38	a	-c	b
39	b	a	-c
40	b	-c	a
41	-c	a	b
42	-c	b	a
43	a	-b	-c
44	a	-c	-b
45	-b	a	-c
46	-b	-c	a
47	-c	a	-b
48	-c	-b	a

Table 9.5: Sequences analyzed for the second layer.

Sample Space No.	Total Space No.	First Sub-pass	Second Sub-pass	Third Sub-pass
1	1	a	b	c
2	2	a	c	b
3	3	b	a	c
4	9	b	-a	c
5	11	c	-a	b
6	28	-b	c	a
7	30	c	-b	a
8	38	a	-c	b
9	41	-c	a	b
10	15	-b	-a	c
11	34	b	-c	-a
12	43	a	-b	-c
13	19	-a	-b	-c
14	20	-a	-c	-b

Table 9.6: The R_2'' -matrix.

	a	b	c	-c	-b	a
a	0	2	2	1	1	0
b	1	0	1	1	0	1
c	1	1	0	0	1	1
-c	1	1	0	0	1	1
-b	1	0	1	2	0	1
-a	0	1	2	1	1	0

Table 9.7: The R'_1 -matrix.

	First Position	Second Position	Third Position
a	4	2	2
b	3	1	4
c	2	2	4
-c	1	3	2
-b	2	3	1
-a	2	3	1

The recipe to train the surrogate model for approximation of the distortion in the girth weld, builds through the priority list as follow:

- First position:

The order is 1.

The starting sub-pass has always the exact answer and therefore this sub-pass of order one has always type 1 occurrence. $R'_1(1^{st})$.

- Second position:

The order is 2.

Search for any type 1 of order 2 for this sequence vector. $R'_2(1^{st}, 2^{nd})$.

If nothing is found, search for any type 2 of order 2, Regardless of the direction of the non-dominant sub-passes only. $R''_2(\pm 1^{st}, 2^{nd})$. Second sub-pass is assumed to be dominant for example. This is because it has been observed that the contribution from the direction is less effective than the position and therefore we search for the highest order of type 1 around the dominant sub-pass regardless of the sign/direction of other sub-passes for the next level of approximation.

If nothing is found, search for any type 2 of order 2, regardless of the direction of any sub-passes. $R''_2(\pm 1^{st}, \pm 2^{nd})$.

If nothing is found, truncate from the left and search for type 1 of order 1. $R'_1(2^{nd})$.

The existence of this term is guaranteed by the R'_1 -matrix.

- Third position:

The order is 3.

Search for any type 1 of order 3 for this sequence vector. $R'_3(1^{st}, 2^{nd}, 3^{rd})$.

If nothing is found, search for any type 2 of order 3, regardless of the direction of the non-dominant sub-passes only. $R''_3(\pm 1^{st}, 2^{nd}, \pm 3^{rd})$.

If nothing is found, search for any type 2 of order 3, regardless of the direction of any sub-passes. $R''_3(\pm 1^{st}, \pm 2^{nd}, \pm 3^{rd})$.

If nothing is found, truncate from the left and search for any type 1 of order 2. $R'_2(2^{nd}, 3^{rd})$.

If nothing is found, search for any type 2 of order 2, regardless of the direction of the non-dominant sub-passes only. $R''_2(2^{nd}, \pm 3^{rd})$.

If nothing is found, search for any type 2 of order 2, regardless of the direction of any sub-passes. $R''_2(\pm 2^{nd}, \pm 3^{rd})$.

If nothing is found, truncation from left and search for any type 1 of order 1. $R'_1(3^{rd})$. The existence of this term is guaranteed by the R'_1 -matrix.

9.6 Surrogate Results and Comparison with Exact Results

Maximum displacement occurs in and around the weld. Three FEM nodes, denoted A, B, and C, are picked on one side of the weld path to implement the surrogate approximation. A is in the mid-path of sub-pass a, B is in the mid-path of sub-pass b, and C is in the mid-path of sub-pass c as shown in Fig. 9.8. The surrogate model for position C was constructed. The procedure is similar for A and B. However the algorithm can be repeated for all nodes in the FEM domain or any selection

of nodes in the domain such as a circle around the weld to construct a surrogate model. Displacements [mm] in X, Y and Z directions are denoted as C1, C2 and C3 respectively.

Fig. 9.9 shows the displacement for all settings in the sample space. Looking at these plots, the behaviour can be classified into 3 main categories. The dominant sub-pass for the position C is the closest sub-pass that is c. The position of sub-pass c was found to classify the behaviour into 3 categories which are c in position 1, 2, or 3. That encouraged the generation of the preference list based on the position of the dominant sub-pass.

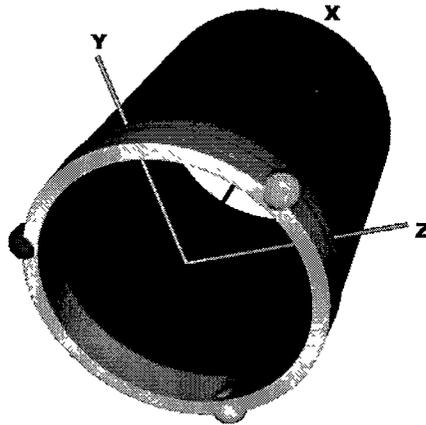


Figure 9.8: Position A (blue), B (green), and C (yellow) picked for constructing the surrogate model. Coordinate axes are shown in red, yellow and green for the X, Y and Z respectively

From the total space, three subsequences were picked out of the sample space to show how to construct the surrogate model and also to show how accurate the surrogate model approximates the displacement in comparison to exact FEM.

The first sequence (c, a, b) has the dominant sub-pass c in the first position. This is number 5 in Table 9.4 for the total space. Table 9.8 shows all possible choices for each position based on the chosen priority list. Priority decreases from the top to

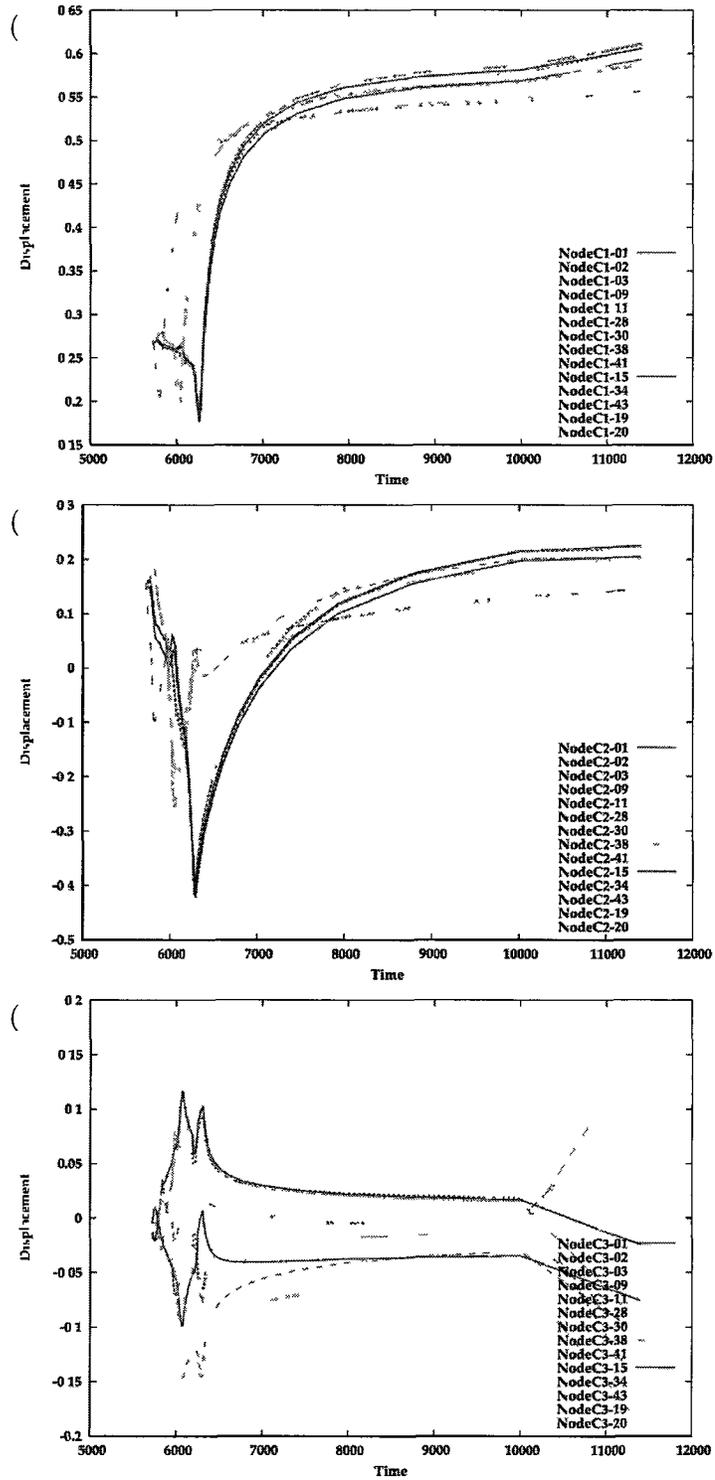


Figure 9.9: Displacement [mm] results for all settings in the sample space. Directions X, Y, and Z are illustrated in 9.9(a), 9.9(b) and 9.9(c)

the bottom of the list because as error is introduced to the surrogate approximation from the top to the bottom of the list. Table 9.6 provides a better sample space to push the available choices toward the top of the list. From the selected sample space, Table 9.5, $R'_1(c)$ was picked for the first position, $R''_2(c, \pm a)$ was picked for the second position and $R''_3(c, \pm a, \pm b)$ was picked for the third position. Therefore the approximation could be achieved by Eq. 9.2.

Table 9.8: List of choices for sequence (c, a, b).

c	a	b
$R'_1(c)$	$R'_2(c, a)$	$R'_3(c, a, b)$
	$R''_2(c, \pm a)$	$R''_3(c, \pm a, \pm b)$
	$R''_2(\pm c, \pm a)$	$R''_3(\pm c, \pm a, \pm b)$
	$R'_1(a)$	$R'_2(a, b)$
		$R''_2(\pm a, \pm b)$
		$R'_1(b)$

$$Approx.(c, a, b) = R'_1(c)_{11} + R''_2(c, \pm a)_{11} + R''_3(c, \pm a, \pm b)_{11} \quad (9.2)$$

Each term comes from one result in the sample space. In this case, all of them come from the number 5 in the sample space or setting number 11 in the total space – subscripts. Fig. 9.10 illustrates this surrogate approximation in comparison with the exact result obtain from the FEM analysis for the sequence (c, a, b). This shows a very good approximation of the behaviour so that the maximum error between the exact solution and the approximate result is less than 0.04 [mm] at the worst case for the Z direction at the end.

Plot legends are described here for convenience; for example, NodeC1-05, corresponds to node C, component 1 (or x), and sequence number 5 in the total space of sequences as per Table 9.4, taken from the FEM analysis. If the legend is followed by terms such as, P1P2P3, the plot is from the surrogate model. For example, Node-C2-28-P2P3, shows the surrogate results for the node C, component 2 or y, sequence number 28 in the total space of sequences for the second and third terms only. In effect, the FEM result is replaced for the time duration of the second and third sub-pass from sequence 28 by the result from surrogate model.

The second sequence (b, c, a) has the dominant sub-pass c in the second position. This is number 4 in Table 9.4 for the total space. Table 9.9 shows all possible choices for each position based on the chosen priority list. From the current sample space, Table 9.5, $R'_1(b)$ was picked for the first position, $R''_2(\pm b, c)$ was picked for the second position and $R''_3(\pm b, c, \pm a)$ was picked for the third position. Therefore the approximation could be achieved by Eq. 9.3.

Table 9.9: List of choices for sequence (b, c, a).

b	c	a
$R'_1(b)$	$R'_2(b, c)$	$R'_3(b, c, a)$
	$R''_2(\pm b, c)$	$R''_3(\pm b, c, \pm a)$
	$R''_2(\pm b, \pm c)$	$R''_3(\pm b, \pm c, \pm a)$
	$R'_1(c)$	$R'_2(c, a)$
		$R''_2(c, \pm a)$
		$R''_2(\pm c, \pm a)$
		$R'_1(a)$

$$Approx.(b, c, a) = R'_1(b)_3 + R''_2(\pm b, c)_{28} + R''_3(\pm b, c, \pm a)_{28} \quad (9.3)$$

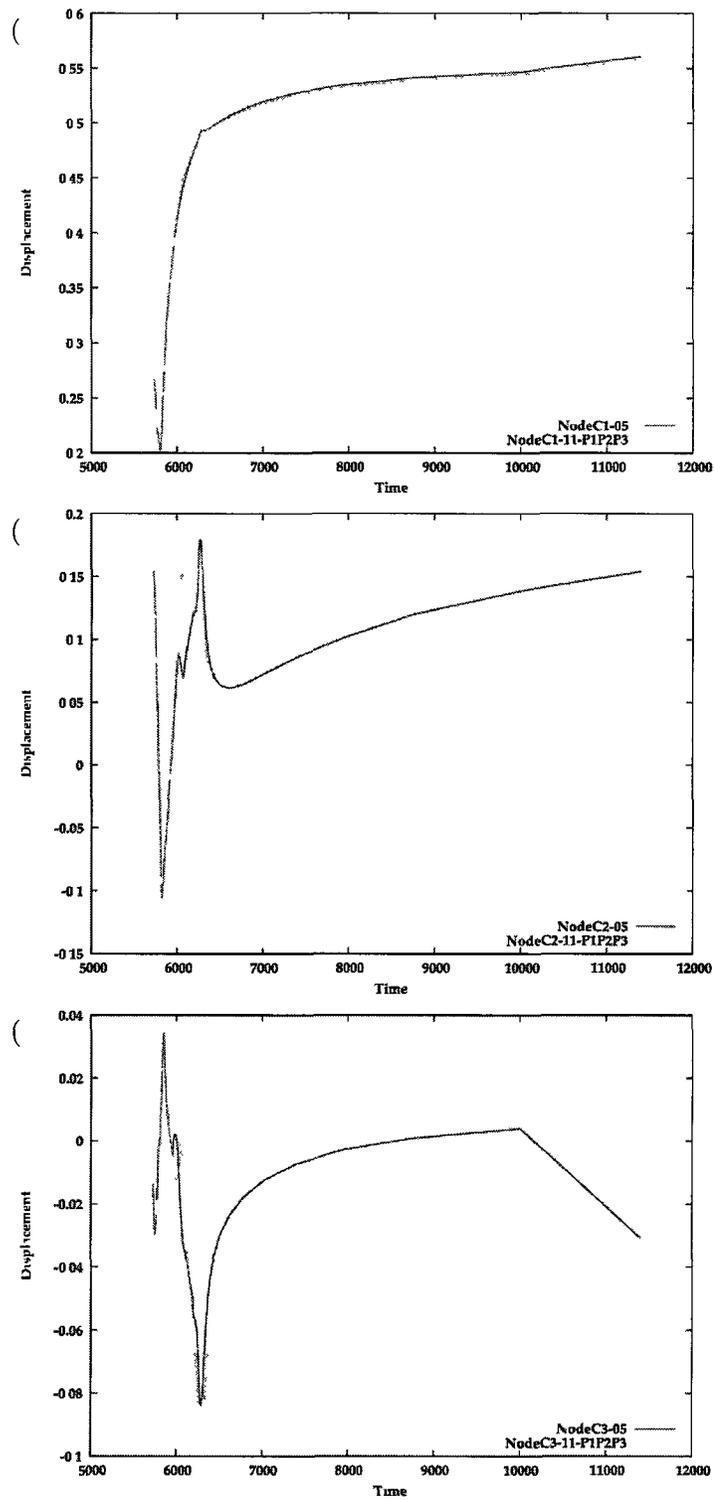


Figure 9.10: Comparison between the displacement [mm] obtained from FEM and surrogate approximation for the sequence (c, a, b). Directions X, Y, and Z are illustrated in 9.10(a), 9.10(b), and 9.10(c).

Each term comes from one result in the sample space. In this case, the first term comes from the number 13 in the sample space or setting number 19 in the total space. The second and third terms come from number 6 in the sample space or setting number 28 in the total space. Fig. 9.11 illustrates this surrogate approximation in comparison with the exact result obtain from the FEM analysis for the sequence (b, c, a). This shows a very good approximation of the behaviour with the exception of the last cooling time step for Z direction.

The third sequence (-a, b, c) has the dominant sub-pass c in the third position. This is number 7 in Table 9.4 for the total space. Table 9.10 shows all possible choices for each position based on the chosen priority list. From the current sample space, Table 9.5, $R'_1(-a)$ was picked for the first position, $R''_2(\pm a, \pm b)$ was picked for the second position and $R''_3(\pm a, \pm b, c)$ was picked for the third position. Therefore the approximation could be achieved by Eq. 9.4.

Table 9.10: List of choices for sequence (-a, b, c).

-a	b	c
$R'_1(-a)$	$R'_2(-a, b)$	$R'_3(-a, b, c)$
	$R''_2(\pm a, \pm b)$	$R''_3(\pm a, \pm b, c)$
	$R'_1(b)$	$R''_3(\pm a, \pm b, \pm c)$
		$R'_2(b, c)$
		$R''_2(\pm b, c)$
		$R''_2(\pm b, \pm c)$
		$R'_1(c)$

$$Approx.(-a, b, c) = R'_1(-a)_{19} + R''_2(\pm a, \pm b)_1 + R''_3(\pm a, \pm b, c)_1 \quad (9.4)$$

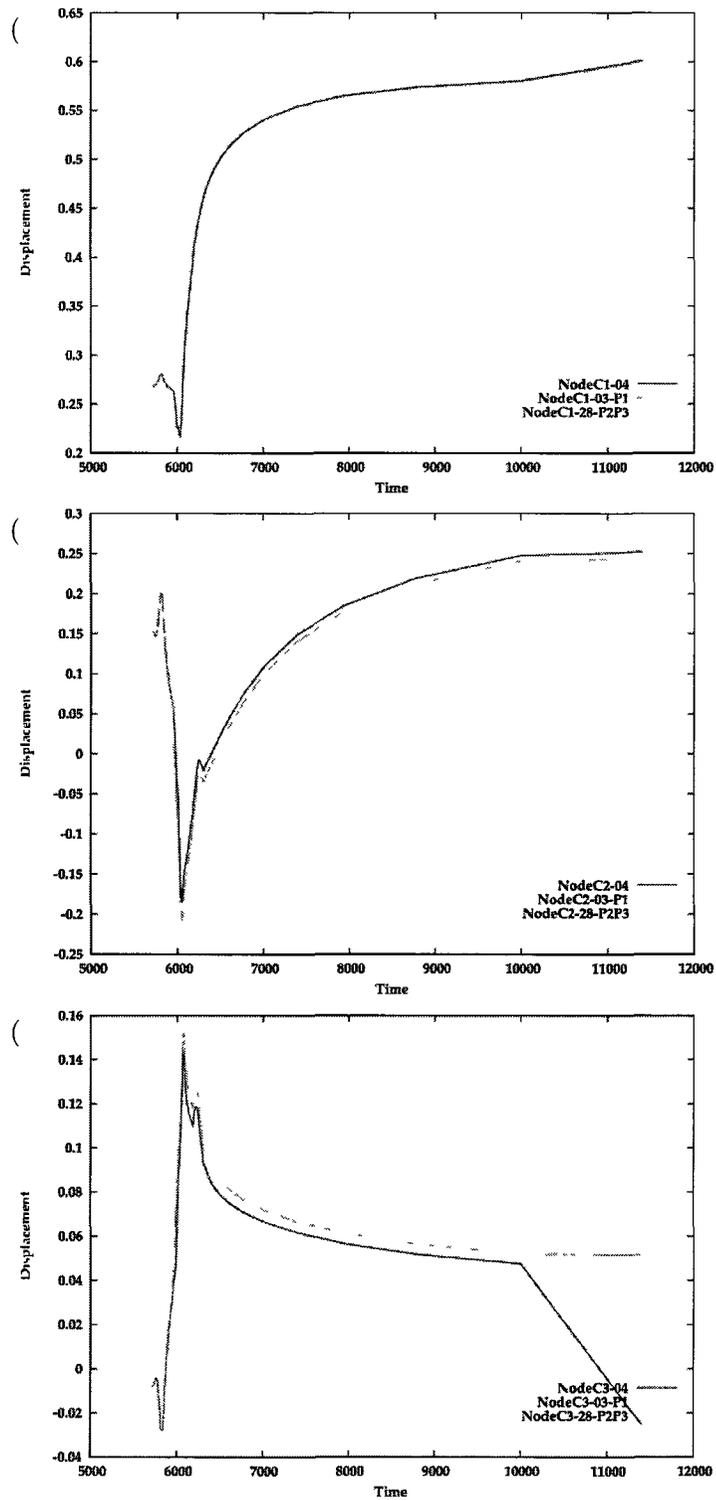


Figure 9.11: Comparison between the displacement [mm] obtained from FEM and surrogate approximation for the sequence (b, c, a). Directions X, Y, and Z are illustrated in 9.11(a), 9.11(b), and 9.11(c).

Each term comes from one result in the sample space. In this case, the first term comes from the number 13 in the sample space or setting number 19 in the total space. The second and third terms come from number 1 in sample space or setting number 1 in the total space. Fig. 9.12 illustrates this surrogate approximation in comparison with the exact result obtain from the FEM analysis for the sequence (-a, b, c). This shows a very good approximation of the behaviour almost identical for X and Y direction. The error is not more than 0.02 [mm] for about 6 cm total displacement.

9.7 Find the Optimum

The optimization objective is to find the best sequence to minimize distortion. The recipe to construct our surrogate model has shown a good approximation of the displacement from the weld. Since the surrogate model is cheaper to calculate, this suggests it to be used for the total space and generate the displacement for all nodes in the total space. Among those results, the sequence (c, b, a), number 6 in Table 9.4, shows the smallest displacement and the sequence (-a, -b, -c), number 19 in Table 9.4, shows the biggest displacement. To check if the algorithm finds the right optimum, a full factorial DOE matrix including all 48 analyses of Table 9.4 ran in the multi-project mode using 4 cores finished in 10 days. The accurate model and the surrogate model both find the best sequence (c, b, a) and the worst sequence (-a, -b, -c).

9.8 Compare the Best and Worst

To compare the results between the best sequence (c, b, a) and the worst sequence (-a, -b, -c), the best/smallest displacement was 0.75 [mm] and the worst/biggest

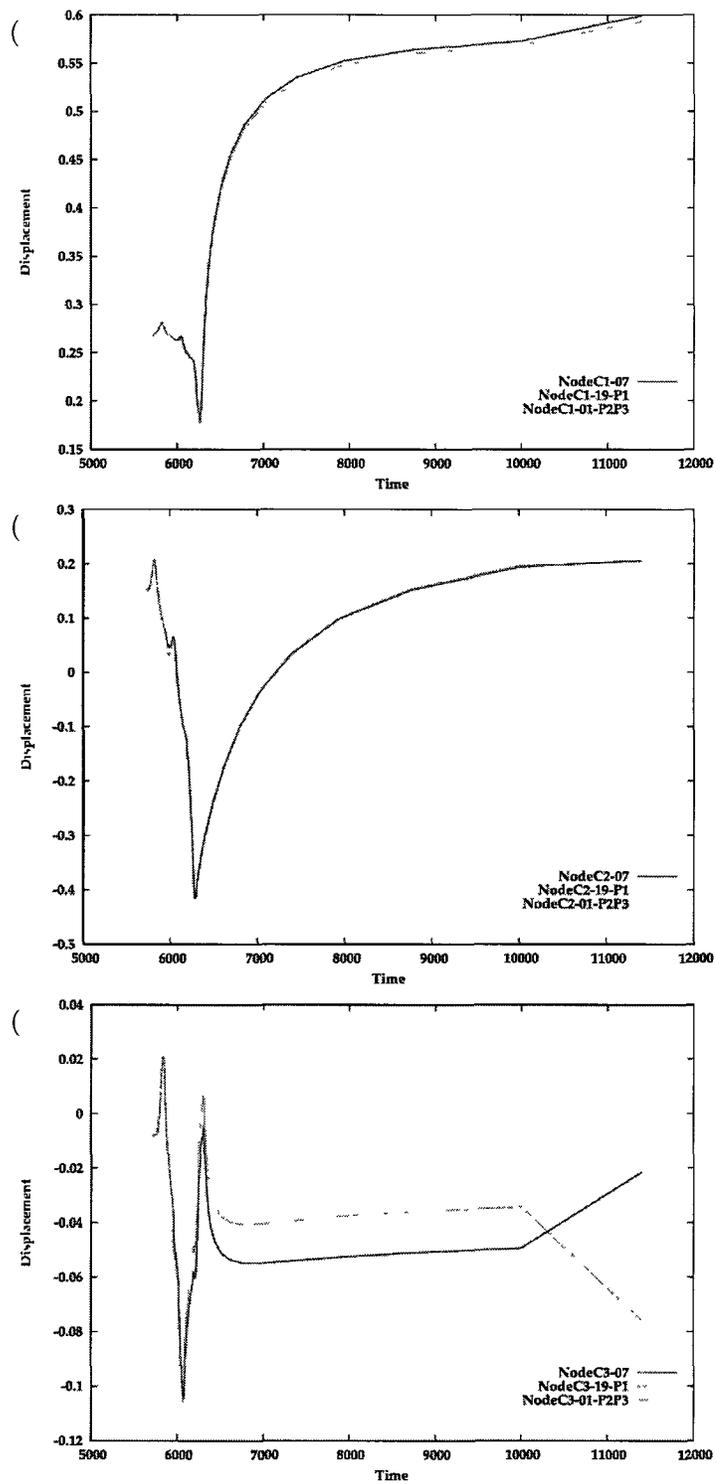


Figure 9.12: Comparison between the displacement [mm] obtained from FEM and surrogate approximation for the sequence (-a, b, c). Directions X, Y, and Z are illustrated in 9.12(a), 9.12(b), and 9.12(c).

displacement was 0.85 [mm] as shown in Fig 9.13. Since the line plots give more quantitative information, The nodes in the pipe's cross section were picked as shown in Fig 9.14(a) and rotated it 360 degrees to generate a mesh illustrated in Fig 9.14(b) and employed it for plotting the results further.

Fig. 9.15 shows the radial displacement plotted for the line number 55 which is -50 degree from the reference line (red). Fig 9.16 shows the radial displacement plotted along the circles number 1, 11, and 14 respectively. Number 1 is the farthest from the weld, number 2 is closest to the weld, and number 12 is in the middle of the weld.

Residual stress, particularly hoop residual stress for pipes, is usually of interest to designers along with the distortion. Fig 9.17 shows the hoop residual stress at the end of the process for the best (a) and the worst (b) sequence. These plots are on axial lines number 8, 10, 12, 1, and 20 where the first three are about 45 degrees from the reference line between the tack welds and the last two are on the first and second tack welds. Fig 9.17(c) compares the line 10 and 20 of the both cases.

Fig 9.18 shows the hoop residual stress at the end of the process for the best (a) and the worst (b) sequence along the circle number 3, 6, 8, 11, and 12 where the number 3 is the farthest and number 12 is the closest to the weld mid line. Fig. 9.18(c) compares hoop residual stress of the line 3 and 12 of the both cases.

9.9 Closure of the Analysis

When it is not feasible to use an accurate model to generate the results for all points in a discrete design space, reduced space methods are used. These methods work on a sample sub-space of the design space instead of the total space. Surrogate models are one kind of reduced space method. They construct an estimator from the information acquired from a small sample sub-space, to approximate the results over the complement space of the sample sub-space. Compared to an accurate model, the

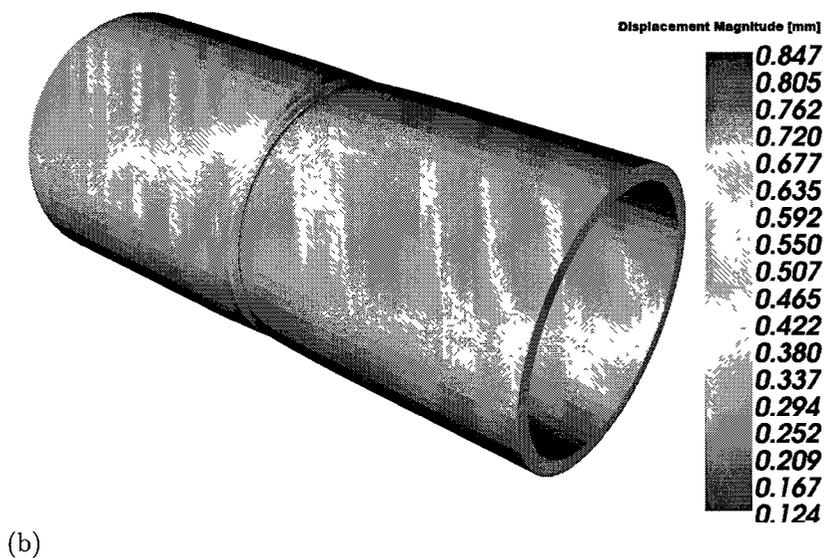
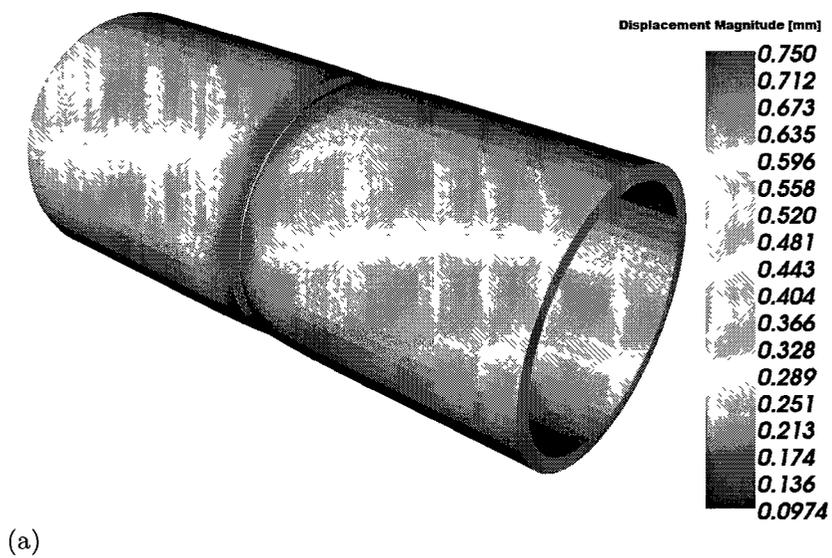
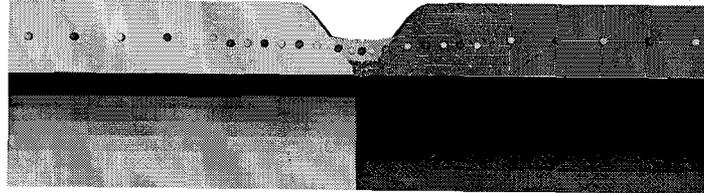
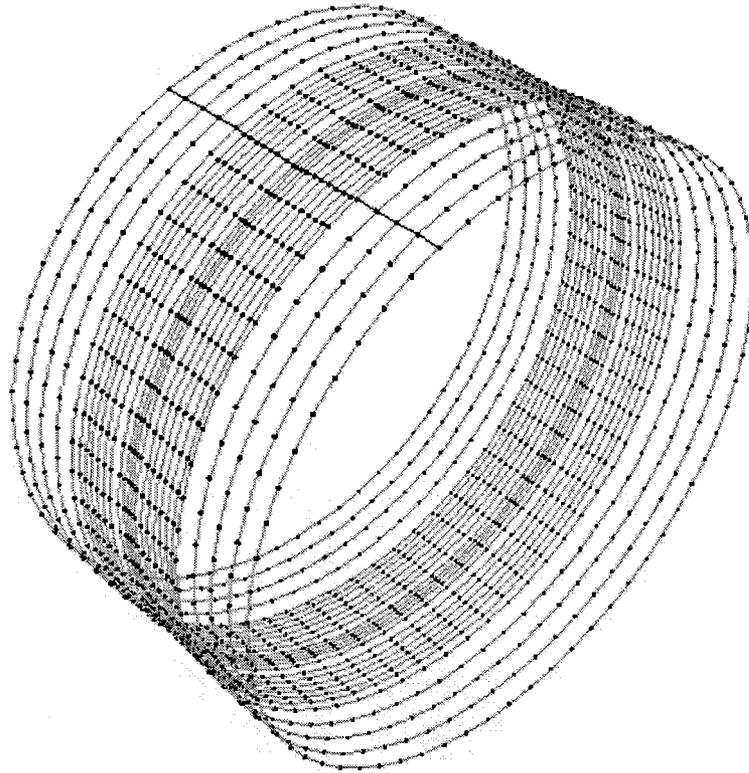


Figure 9.13: Comparison between the displacement [mm] for the best sequence (c, b, a) 9.13(a) and the worst sequence (-a, -b, -c) 9.13(b). Wall thickness is 17.5 [mm].



(a)



(b)

Figure 9.14: Mesh employed for plotting the results. The red line is a reference line.

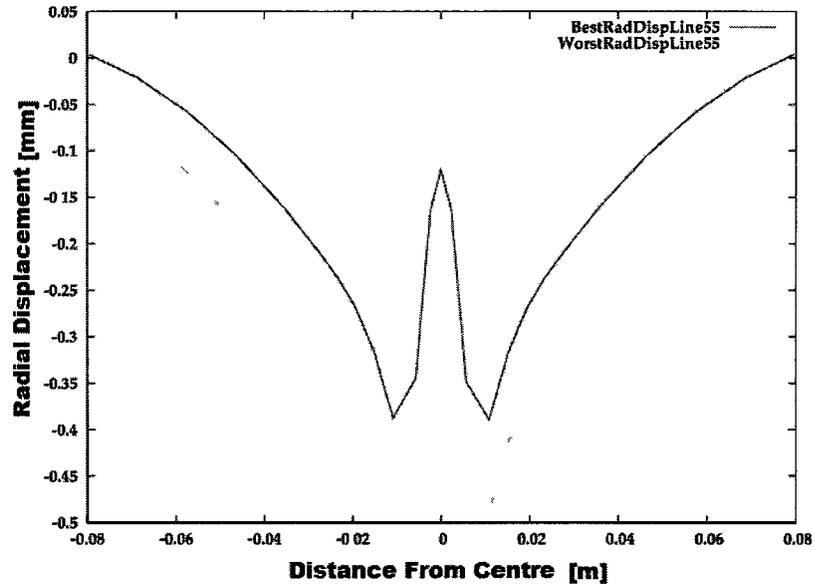


Figure 9.15: Radial displacement plotted for the line number 55 which is -50 degree from the reference line. Red shows the best sequence and the green shows the worst sequence.

surrogate models are significantly cheaper and therefore it could be feasible to use them over tens of thousands of points in design space. Although surrogate approaches can be efficient, a concern is that users have to invent the recipe for each specific case and there is no general algorithm. Each model/recipe might work properly in some cases but not for all cases.

In this analysis, a surrogate model is demonstrated to minimize the distortion in a girth weld of a pipe welded with six sub-passes by analyzing only 14 weld sequences from space of total sequences of welding. A comparison between the results achieved by the surrogate model and the exact FEM analysis shows the accuracy of the algorithm. The best and worst sequence are analyzed to compare the radial displacement and hoop residual stress as two important criteria for designers.

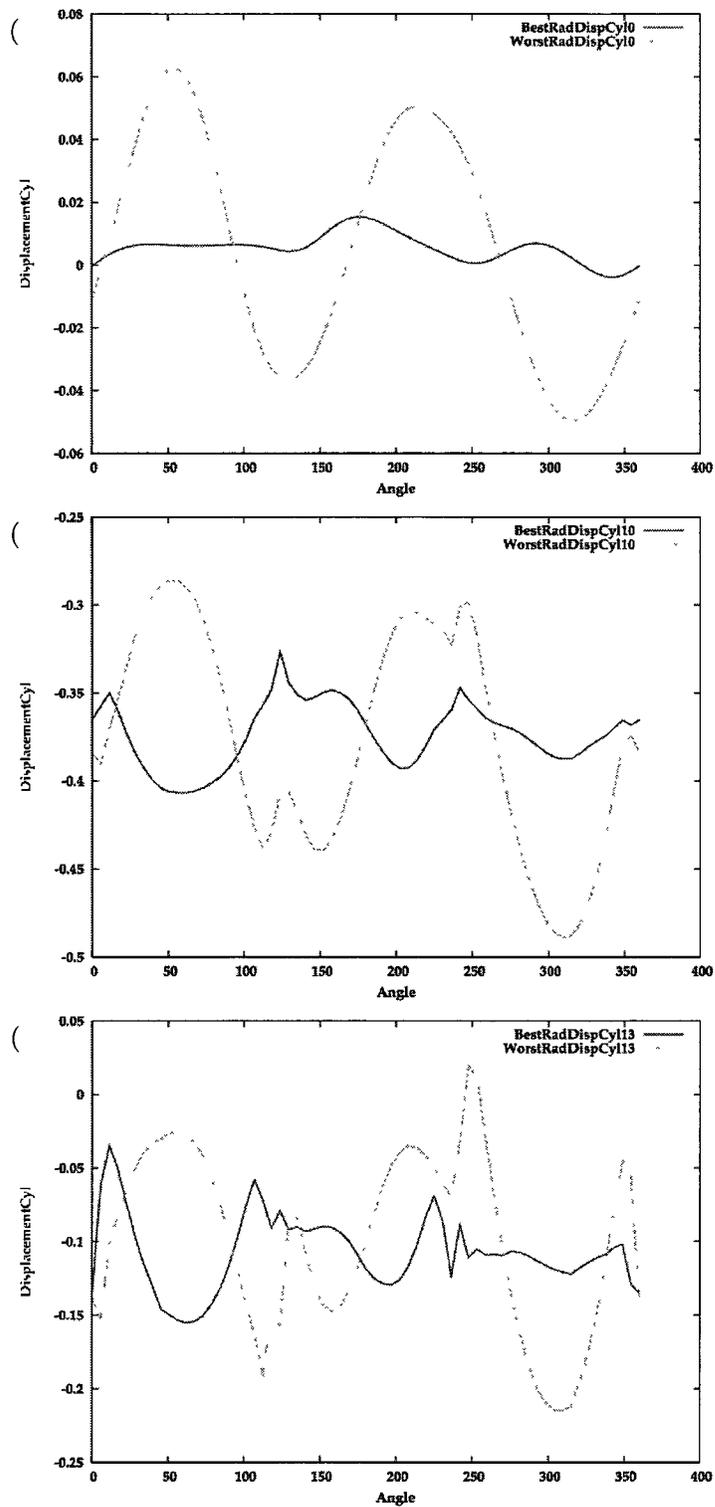


Figure 9.16: Radial displacement plotted along the circles number 1, 11, and 14 respectively. Number 1 is the farthest from the weld, number 2 is close to the weld and number 12 is in the middle of the weld. Red shows the best sequence and the green shows the worst sequence.

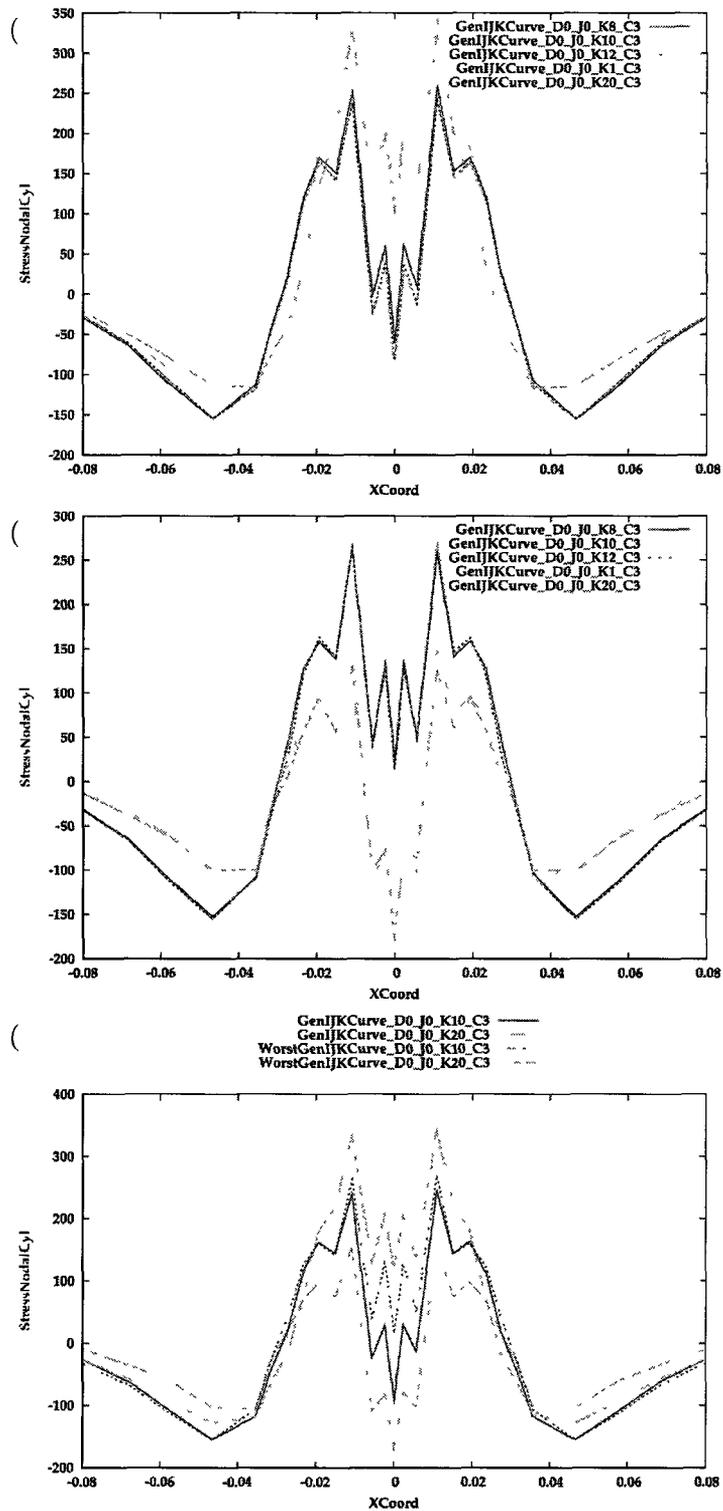


Figure 9.17: Hoop residual stress at the end of the process for the best (a) and the worst (b) sequence. These plots are on axial lines number 8, 10, 12, 1, and 20 where the first three are about 45 degrees from the reference line between the tack welds and the last two are on the first and second tack welds. Fig. 9.17(c) compares the line 10 and 20 of both cases.

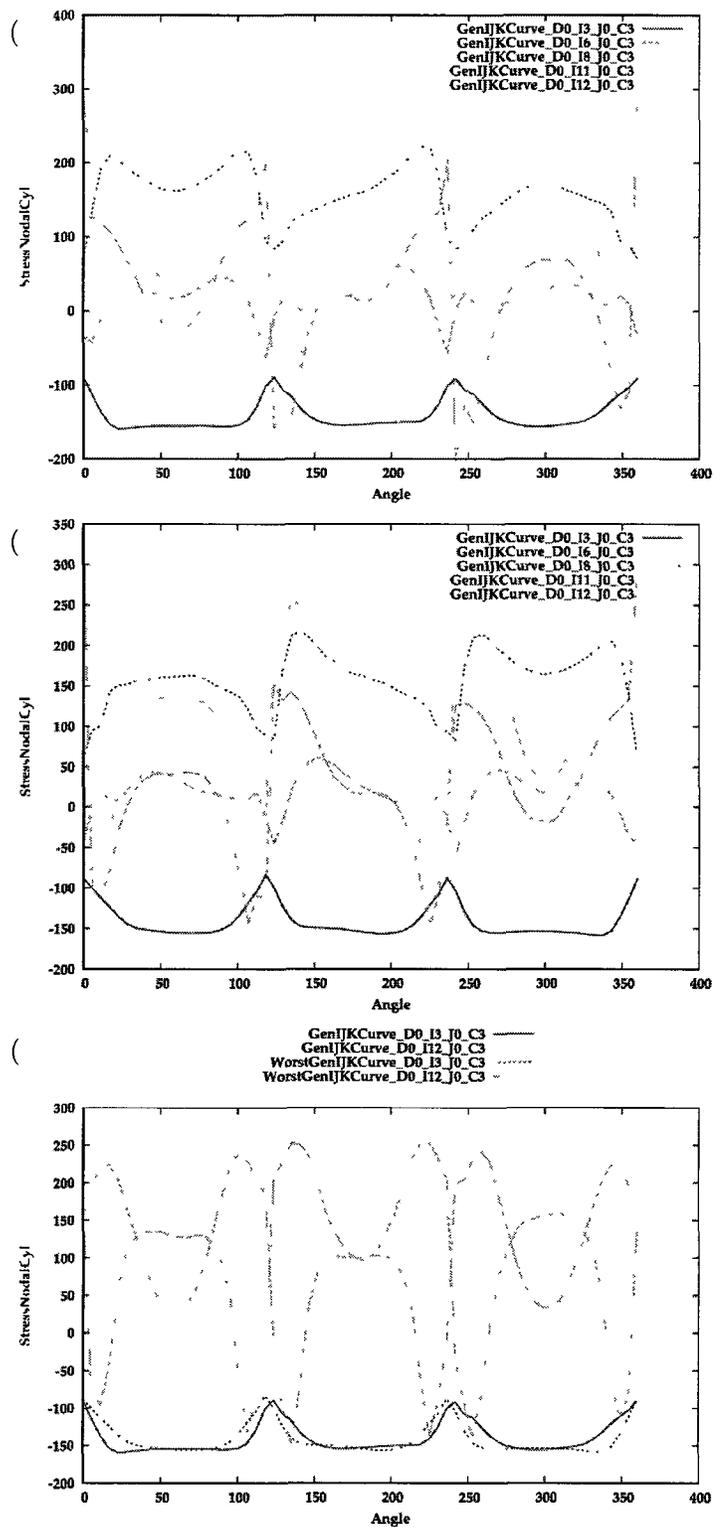


Figure 9.18: Hoop residual stress at the end of the process for the best (a) and the worst (b) sequence along the circle number 3, 6, 8, 11, and 12 where the number 3 is the farthest and number 12 is the closest to the weld mid line. Fig. 9.18(c) compares hoop residual stress of the line 3 and 12 of the both cases.

Chapter 10

Designer-Driven Exploration of CWM Design Space

This chapter compares the analyses presented in chapters 5, 6, 7, 8 and 9 in terms of DOE matrices used and CPU time.

10.1 Problem Background

The usual classical approach to solve an optimization or a design problem, would be to design a Design-Of-Experiment (DOE) test matrix and perform the physical experiments defined by the test matrix. This is expensive in time and money. Because the DOE test matrix is designed with limited information and uncomplete understanding or insight of the design space, it may not be the optimal DOE matrix. Given such a proposed DOE test matrix, there can be significant benefits from running a multiple computer simulation to evaluate all of the points in the design space that are defined by the DOE test matrix. This capability gives designers an opportunity to construct more efficient DOE matrices. For the welding industry, computer models for DOE matrices and CWM optimization are only now becoming feasible because of improvements in software, numerical algorithms and computer hardware.

Today's welding technology has largely been developed based on theory and physical experiment. However, computational models and numerical experiments have contributed little to the development of modern welding technology. Computer models of welding processes began to be developed in the early 1980s and are now reaching a degree of maturity. The majority of papers published on computational models of welds do only one numerical analysis, and a small fraction of the published papers describe more than one analysis.

10.2 Optimization Problems and Comparison

Four continuous optimization problems were solved in chapters 5, 6, 7, and 8. A discontinuous combinatorial optimization problem was solved in chapter 9. Table 10.1 summarized the DOE matrix sizes, total CPU times, and number of cores used in the project discussed in these chapters. In all cases, the machine was a 3.3 GHz Intel quad-core processor.

The human time to set up a CWM optimization problem is mostly the time to set up the analysis of the first base or reference design and the user-decision-time to design the DOE matrix. Once a design is created, most of the human time is spent in designing the DOE matrices.

In some cases, a DOE matrix can be reused for different projects. For example, the DOE matrix needed to do the combinatorial CWM optimization to find the sequence of sub-passes of a weld pass divided into N sub-passes.

An expert in CWM setup can create a base CWM design and ship it to the design team. They can implement the DOE matrix to do designer-driven CWM optimization that could require hundreds of analyses. The design team does not need to know how to setup the CWM base project. The design team can focus on design space exploration. This is similar to solving a math problem using Mathematica

Table 10.1: DOE matrix sizes, total CPU times, and number of cores used in the project discussed in chapters 5, 6, 7, 8 and 9.

Project	Analysis	Single Run	DOE Matrix Size	Total CPU Time	Num of Cores
Space Discret	Nodal Prebend	48 m	45×2	9 6 h	4
Space Discret	Parab Prebend	48 m	44×2	8 8 h	4
Direct Search	Taguchi	48 m	9×4	2 4 h	4
Direct Search	R-and- η	48 m	36×4	7 2 h	4
Direct Search	Regular Search	48 m	Table 6 5	7 2 h	4
Direct Search	Least Sq Search	48 m	Table 6 6	5 6 h	4
Regression Model	O F A T	13 m	30×1	1 73 h	4
Regression Model	Regression	13 m	81×4	4 55 h	4
Hot Cracking	D Ellip Control	0 6 m	9 × (13×4)	0 37 h	4
Hot Cracking	Sensitivity	5 6 m	28×4	0 67 h	4
Surrogate Model	Layer 1	15 h	6×3	30 h	3
Surrogate Model	Surrogate Layer 2	20 h	14×3	80 h	4
Surrogate Model	Full Space	20 h	48×3	240 h	4
Total of 505 CWM Analyses			Total of 33 DOE Matrices	398 12 h	

that enables the user to focus on the problem independent of the mathematical complexity hidden behind the Mathematica commands.

In many cases, a base project could be reused for different DOE matrices. For example, CWM optimization can be implemented for different mitigation techniques to find the best optimum among the techniques. The side-heater and prebending projects discussed in previous chapters for an edge welded bar are examples of such cases.

It is also possible to use ideas from computer learning to generate a lower dimensional function space, such as surrogate models, to reduce the CPU time to analyze a DOE matrix. This is expected to be more important in larger projects where many analyses have to be done or when the CPU time for a single analysis is too long.

Although the structures analyzed in this thesis are relatively simple this methodology is directly applicable to any welded structure. CPU time per analysis would increase for larger more complex structures but human time to set up the DOE matrix

would change little.

The results presented have demonstrated that CWM analysis using a DOE matrix in CWM is now practical for optimizing many of the decisions required in the design of welded structures.

Chapter 11

Conclusion

Developing an automated framework that supports multiple analyses of welded structures saves user time by avoiding repetitive tasks and organizing the evaluation. This enables the user to devote more time to designing a more efficient DOE matrix. The core competency of the framework is an automated implementation of $m \times n$ DOE matrices of a CWM problem without accumulated people time to make multiple setups in addition to an allocation manager for utilizing an efficient CPU time to run. In this framework the user time to setup CWM DOE matrices changes little from a small and simple project to a large and complex project. However, because CPU time is expected to be critical in large and complex projects, reducing the number of analyses in DOE matrices, is important and needs a proper and effective DOE matrix design.

The users' creativity, experience, and skill play a critical role in such a DOE matrix design. Parametric-design DOE matrix can be a proper viewpoint for design in this framework that takes advantage of simultaneous multiple trial solutions versus single-design DOE matrix for one result at a time. However, supporting interactive DOE matrix design and analyses is found to be an efficient way for solving large industrial CWM DOE matrix projects.

The results presented here have demonstrated that DOE matrix in CWM is now practical for optimizing many decisions in the design of welded industrial structures. This is a powerful tool for a designer-driven optimization that a design group is able to do the optimization in the design stage with a knowledge of down stream welding and production engineering which is in contrast with the traditional practice of the designer waiting for the feedback from welding and production engineers to complete the design and handing off the optimization to a specialist. The conjecture of this thesis is that designer-driven optimization of the design of welded structures is now feasible for routine engineering in industry.

11.1 Looking to the Future

A possible short term improvement on the current frame work could be to implement support for perturbing a mesh and its BCs. A longer term improvements could include: a) adding computer-learning machinery to support smarter DOE matrix design, b) using advanced clustering and parallel computing algorithms and c) higher automation in defining and evaluation of the objective function to support a broader range of user-defined functions.

The largest part of the user time for setup of a base project is devoted to mesh the structure and weld filler metals. A significant time saving could be an automated mesher for a welded structures.

Real-time control application of CWM could be one of the main future goals. This requires a computation time to be less than real-time of welding. Coupling this machinery with importing and processing sets of measured data over time, would give a powerful tool for a real-time control of welding that has been weak so far. Automation could save time by avoiding repetitive tasks in importing and processing the data for a real-time weld control. This also requires a code to implement sensitivity analysis

to find/approximate the gradient of an objective function wrt control variables and to generate a correction for control vectors to match predefined key frames.

A continuous improvement is to solve more actual CWM optimization and control problems especially those with a larger design space. Solving such problems, designing and implementing DOE matrices to explore a larger design space is challenging. This challenge is a driving-force to recognize opportunities to improve and enhance the framework. The current framework is practical and affordable for CWM applications with a limited number of design parameters. What other support is required for a new problem is a question that depends on the problem itself.

List of References

- [1] J. Vitek, Models are Great - So What, Adams Lecture, American Welding Society, Atlanta, Georgia, USA, October 30, 2006.
- [2] V. Adams, A. Askenazi, Building Better Products with Finite Element Analysis, OnWord Press 1st ed, Book, ISBN-13: 978-15669016041999, 1999.
- [3] J. F. Lancaster, The Physics of Welding (Materials Science & Technology Monographs), Pergamon 2 ed, Book, ISBN-13: 978-008034076019861986, 1986.
- [4] E. F. Camacho, C. Bordons. Model predictive control, Advanced textbooks in control and signal processing, Second edition, ISBN 1-85233-694-3, Springer Verlag, 2004.
- [5] Garcia, C; Prett, Morari, Model predictive control: theory and practice, Automatica 25, pp.335348. doi:10.1016/0005-1098(89)90002-2, 1989.
- [6] J. Song, J. Y. Shanghvi, P. Michaleris, Sensitivity analysis and optimization of thermo-elasto-plastic processes with applications to welding side heater design, Computer methods in applied mechanics and engineering, Volume 193, pp. 4541-4566, DOI 10.1016/j.cma.2004.03.007, 2004.
- [7] I. Voutchkov, A. Keane, A. Bhaskar, T. Olsen, Weld sequence optimization: The use of surrogate models for solving sequential combinatorial problems, Computer Methods in Applied Mechanics and Engineering, Volume 194, Issues 30-33, pp. 3535-3551, August 2005.
- [8] C. L. Tsai, S. C. Park, W. T. Cheng, Welding Distortion of a Thin-Plate Panel Structure, Welding Research Supplement, pp. 156-s - 165-s, 1999.
- [9] N. O. Okerblom, The Calculation of Deformations of Welded Metal Structures, Translated from Russian, 1958, NRC library, Dept of Scientific and Industrial Research, London 1958.

- [10] V. A. Vinokurov, *Welding Stresses and Distortion: Determination and Elimination*, The British Library Board 1977, Translated by J. E. Baker from Russian, 1968.
- [11] Y. P. Yang, F. W. Brust, Z. Cao, Y. Dong, A. Nanjundan, *Welding-induced distortion control techniques in heavy industries*, 6th International Conference: Trends in Welding Research; Pine Mountain, GA; USA; 15-19 Apr. 2002. pp. 844-849. 2003.
- [12] E. Hughes, *The Adjoint Method for Derivatives*, Personal communication with J. A. Goldak, Ottawa, Nov 2009.
- [13] B. Nicolas, *Topological vector spaces*, Elements of mathematics, Berlin: Springer-Verlag, Book, ISBN 978-3540136279, 1987.
- [14] A. McNamara, A. Treuille, Z. Popovic, J. Stam, *Fluid Control Using the Adjoint Method*, *Journal ACM Transactions on Graphics (TOG)*, Proceedings of ACM SIGGRAPH, Volume 23 Issue 3, doi 10.1145/1015706.1015744 August 2004.
- [15] R. M. Lewis, V. Torczon, M W. Trosset, *Direct Search Method: Then and Now*, *J. Comput. Appl. Math.*, Volume 24, pp. 191-207, 2000.
- [16] M. W. Trosset, *I know it when I see it: Toward a denition of direct search methods*, *SIAG/OPT Views-and-News: A Forum for the SIAM Activity Group on Optimization*, Volume 9, 1997.
- [17] W. H. Press, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery, *Numerical Recipes in C : The Art of Scientific Computing*, Second edition, ISBN 0-521-43108-5, 1992.
- [18] M. B. Giles, N. A. Pierce, *An Introduction to the Adjoint Approach to Design*, *Flow Turbulance and Combustion 65*: 393-415, Kulwer Academic Publishers, 2000.
- [19] G. Taguchi, *Introduction to Quality Engineering*, Asian Productivity Organization, Eighth Symposium on Taguchi Methods, American Supplier Institute, Dearborn, MI, USA, October 1990.
- [20] O. Kempthorne, *The Design and Analysis of Experiments*, Book, 19520101975, 631 pp., 1952.
- [21] Goldak Technology Inc, [http : //www.goldaktec.com/vrweld.html](http://www.goldaktec.com/vrweld.html), March 2011.

- [22] ASME V &V 10-2006, Guide for Verification and Validation in Computational Solid Mechanics, An American National Standard, The American Society of Mechanical Engineers, 2006.
- [23] J. A. Goldak, A. Chakravarti, M. J. Bibby, A New Finite Element Model for Welding Heat Sources, Trans. AIME., Volume 15B, pp. 299-305, June 1984.
- [24] J. A. Goldak, B. Patel, M. J. Bibby, J. E. Moore, Computational Weld Mechanics, Invited Opening Paper for AGARD Workshop - Structures and Materials 61st Panel Meeting, Oberammergau, Germany, pp 1-1 1-32, September 8-13, 1985.
- [25] J. A. Goldak, V. Breiguine, N. Dai, Computational Weld Mechanics: A Progress Report on Ten Grand Challenges, 4-th International Conference on Trends in Welding Research, Gatlinburg, Tennessee, USA, June 5-8, 1995.
- [26] J. A. Goldak, Distortion and Residual Stress in Welded Structures; The Next Generation, 8th International Trends in Welding Research, Callaway Gardens Resort, Pine Mountain, Georgia, USA, June 2-5, 2008.
- [27] K. Masabuchi, Analysis of Welded Structures, Sec. Transient Thermal Stress, Pergamon Press, pp. 172 - 187, Oxford, 1980.
- [28] J. A. Goldak, CR2008-283, Thermo-mechanical modelling of welding induced strains: Numerical validation of the weld build-up process, DRDC Atlantic CR 2008-283, 2008.
- [29] M. Asadi, J. Goldak, J. Nielsen, J. Zhou, S. Tchernov, D. Downey, Analysis of Predicted Residual Stress in a Weld and Comparison With Experimental Data Using Regression Model, International Journal of Mechanics and Materials in Design, Springer Netherlands, DOI 10.1007/s10999-009-9107-5, ISSN 1569-1713 (Print) 1573-8841 (Online), Volume 5, Number 4, Pages 353-364, December 2009.
- [30] A. Mashaie, Error Estimates for Finite-Element Solutions of Elliptic Boundary Value Problems, PhD thesis, Supervised J. A. Goldak, Mechanical and Aerospace Engineering Dept. Carleton University, Ottawa, Canada, 1990.
- [31] The Lincoln Electric Co. *http : //www.lincolnelectric.com/ knowledge/articles/ content/ distortion.asp*, April 2010.
- [32] D. Radaj, Welding residual stresses and distortion; Calculation and measurement, Revised edition, DVS Verlag, ISBN 3-87155-791-9, 2002.

- [33] V. Torczon, On the convergence of the multidirectional search algorithm, SIAM J. Optim., Volume 1, pp. 123-145, 1991.
- [34] J. Rotemberg, M. Woodford, An Optimization-Based Econometric Framework for the Evaluation of Monetary Policy, NBER Macroeconomics Annual 12, pp. 297-346, 1997.
- [35] T. G. Kolda, R. M. Lewis, V. Torczon, Optimization by Direct Search: New Perspectives on Some Classical and Modern Methods, SIAM REVIEW, Vol. 45, No. 3, pp. 385482, DOI. 10.1137/S0036144502428893, 2003.
- [36] R. Unal, E. B. Dean , Taguchi Approach to Design Optimization for Quality and Cost: An Overview, Annual Conference of the International Society of Parametric Analysts, 1991.
- [37] ReliaSoft Corp. Online Course on DOE++ 2008, *http* : [//www.weibull.com/DOEWeb/limitations_of_taguchis_approach.htm](http://www.weibull.com/DOEWeb/limitations_of_taguchis_approach.htm), accessed September 2010.
- [38] A. Paradowska, J. W. H. Price, R. Ibrahim, Trevor Finlayson, A neutron diffraction study of residual stress due to welding, Journal of Materials Processing, 164-165, pp. 1099-1105, 2005.
- [39] B. A. B. Anderson, Journal of Engineering Materials and Technology, Trans. ASME, Vol. 100, 356-362, 1978.
- [40] M. Gu, J. Goldak, Steady State Formulation for Stress and Distortion of Welds, Journal of Engineering for Industry, pp. 467-474, Volume 116, November 1994.
- [41] B. F. Iron, S. Ahmad, Techniques of Finite Elements, Ellis-Horwood Limited, ISBN 0-85312-130-3, 1980.
- [42] A. Pahkamaa, J. Pavasson, A new Welding Modeling Approach in Simulation Driven Design, M. Sc. Thesis, Mechanical Eng., Lulea University of Technology, Dept. of Applied Physics and Mechanical Engineering, Division of Computer Aided Design, ISSN: 1402-1617, 2009.
- [43] J. Pan, Arc Welding Control, Book published by Woodhead Publishing, ISBN 185573687X, 2003.
- [44] T. Zacharia, Dynamic Stress in Weld Metal Hot Cracking, Welding Research Supplement 164-172-s, July 1994.

- [45] C. E. Cross, N. Coniglio, Weld Solidification Cracking: Critical Conditions for Crack Initiation and Growth. Hot Cracking Phenomena in Welds II, Springer, pp. 47-66, 2008.
- [46] [4] Y. P. Yang, P. Dong, J. Zhang, X. Tian, A Hot-Cracking Mitigation Technique for Welding High-Strength Aluminum Alloy, WJ New York, 2000.
- [47] F. Matsuda, H. Nakagawa, K Nakata, H. Kohmoto, Y. Honda, Quantitative Evaluation of Solidification Brittleness of Weld Metal During Solidification by In-Situ Observation and Measurment (RepOrt I) - Development of the MISO Technique, Transaction of JWRI, Welding Research Institute of Osaka University, 1983.
- [48] F. Matsuda, H. Nakagawa, H. Kohmoto, Y. Honda, Y. Matsubara, Quantitative Evaluation of Solidification Brittleness of Weld Metal During Solidification by In-Situ Observation and Measurment (RepOrt II) - Solidification Ductility Curves for Steels with the MISO Technique, Transaction of JWRI, Welding Research Institute of Osaka University, 1983.
- [49] A. Pahkamaa, L. Karlsson, J. Pavasson, M. Karlberg, M. Nasstrom, J. Goldak, A Method to Improve Efficiency in Welding Simulation for Simulation Driven Design, International Design Engineering Technical Conference & Computers and Information in Engineering, ASME Conference, DETC2010-28664, August 15-18, Montreal, Quebec, Canada, 2010.
- [50] L. Karlsson, A. Pahkamaa, M. Karlberg, M. Lofstrand, J. Goldak, J. Pavasson, Mechanics of Materials and Structures - a Simulation-Driven Design Approach, Under review to be published. 2011.
- [51] D. R. Jones, A Taxonomy of Global Optimization Methods Based on Response Surfaces," Journal of Global Optimization, 21:345-383, 2001.
- [52] C. T. Kelley, Iterative Methods for Optimization, Book, Frontiers in applied mathematics, ISBN 0-89871-433-8, SIAM, 1999.
- [53] D. Jones, M. Schonlau, W. Welch, Efficient global optimization of expensive black box functions, J. Global Optim. Volume 13, pp. 455-492, 1998.
- [54] M. Gu, J. A, Goldak, Steady State Thermal Analysis of Welds with Filler Metal Addition, Can Metal. Q., Volume 32, pp. 49-55, 1993.
- [55] O. Zienkiewicz, R. Taylor, The Finite Element Method, McGraw-Hill 4th ed., Volume 2, 1989.

- [56] R. Kumar, M. Coulombe, S. Tchernov, J. A. Goldak, E. Johnson, M. El-Zein, A Model Equation for the Convection Coefficient for Thermal Analysis of Welding Structures, 8th International Trends in Welding Research, Callaway Gardens Resort, Pine Mountain, Georgia, USA, June 2008.
- [57] J. D. Francis, Welding Simulations of Aluminum Alloy Joints by Finite Element Analysis, M. Sc. Thesis Virginia Polytechnic Institute and State University, April 2002.
- [58] J. A. Goldak, M. Akhlaghi, Computational Welding Mechanics, Springer, ISBN-13 978-0387-23287-4, e-ISBN 0-387-23288-5, 2005.
- [59] K. Nishiguchi, T. Ohji, H. Matsui, Study on Bead Surface Profile, IIW Doc 212-391-77, 1977.
- [60] T. Ohji, A. Ohkubo, K. Nishiguchi, Mathematical Modelling of Molten Pool in Arc Welding, IUTAM Symposium, Lulea, Sweden 1991.
- [61] T. Ohji, A. Ohkubo, K. Nishiguchi, Mathematical Modeling of Molten Pool in Arc Welding, Mechanical Effects of Welding, pp 207-214, Publ. Springer, Berlin, 1992.
- [62] D. F. Watt, L. Coon, M. J. Bibby, J. A. Goldak, C. Henwood, Modelling Microstructural Development in Weld Heat-Affected Zones, (Part A), Acta Met., Volume 36, No. 11, pp. 3029-3035, 1988 .
- [63] C. Henwood, M. J. Bibby, J. A. Goldak, D. F. Watt, Coupled Transient Heat Transfer-Microstructure Weld Computations (Part B), Acta Met., Volume 36, No. 11, pp. 3037-3046, 1988.
- [64] J. Kirkaldy, D. Venugoplan, Prediction of Micro-structure and Hardenability in Low Alloy Steels, Phase Transformation in Ferrous Alloys, Proceedings of International Conference; Trends in Welding Research, Pine Mountain, Georgia USA, June 1-6 2008.
- [65] C. H. Gur, J. Pan, Handbook of Thermal Process Modelling of Steels, CRC Press 2008.
- [66] M. F. Ashby, K.E. Easterling, A First Report on Diagrams for Grain Growth in the Welds, Acta Metall., 82 11 1969-1978, 1982.

- [67] J. C. Simo, Numerical Analysis of Classical Plasticity, Handbook for Numerical Analysis, Volume IV, ed. by P. G. Ciarlet, J. J. Lions, Elsevier, Amsterdam, 1998.
- [68] M. Asadi, J. A. Goldak, Mathematical Context of the Adjoint Method, Internal report to J. A. Goldak, Carleton University, Ottawa, June 2010.
- [69] R. L. Haupt, S. E. Haupt, Practical Genetic Algorithms, Second Edition, A John Wiley & Sons Inc, ISBN 0-471-45565-2, 2004.
- [70] D. Byatt, I. Coope, C. Price, 40 Years of the Nelder-Mead Algorithm, Seminar note from University of Canterbury, New Zealand, 2003.

Appendix A

Multi-Physics of Welding

CWM is a multi-physics problem. A full computational model of a welding process includes transient thermal, evolution of microstructure and stress/strain analysis. These fields are coupled such that the simulation starts by computing the transient temperature distribution, then evaluating the microstructure from the transient thermal profiles and then computing the stress state including thermal expansion and phase transformation using the thermal and microstructure results. This coupling might exchange data between the time steps such as the effect of the latent heats from phase transformations that act as a heat sink or source in the thermal computation or the effect of mechanical deformation on the formation of martensite/bainite. The basic governing equations for thermal and stress fields are the conservation of mass, energy and momentum. Microstructural analysis is based on an algorithm developed for the material under analysis and could use, for example, the pseudo-Iron-Carbon phase diagram for low-alloy steels or a grain nucleation, growth and coarsening model for aluminum alloys. A graphical user interface (GUI) provides different options to be chosen among the microstructure algorithms and the user picks the appropriate one to be coupled with the thermal and stress solver.

A.1 Thermal Analysis in Welding

A weld applies a moving local high intensity power source to the part that generates a sharp thermal profile in the weld pool, heat affected zone (HAZ) and around the weld. This causes a local change in microstructure, stress and strain state that can have a large effect on the whole structure. The 3D transient temperature is basically determined by solving the partial differential equation for the conservation of energy which is given in Eq. A.1 for a Lagrangian or material formulation.

$$\dot{h} + \nabla \cdot (-\kappa \nabla T) = Q \quad (\text{A.1})$$

Because the heat source is moving we can define an Eulerian frame with origin at the center of the heat source. Using an Eulerian frame enables longer time steps for the weld when the analysis approaches steady state. But the Eulerian frame introduces an advection term as in Eq. A.2 into the FEM equation. This term is nonlinear and hard to solve. Gu [54] implemented a weld model in the Eulerian frame. This is basically an Eulerian mesh fixed in space tied to the arc and the part moves under the arc. CWM analyses usually use a material frame for modeling where the heat source moves in discretized time as steps in series of spot welds. This frame requires sufficiently small time steps to approximate a continuous heat source. The transient heat equation can be solved by the standard Lagrangian finite element method [55]. It maps the Eulerian thermal field into the material mesh by tracking along the flow lines for each time step. Because the elements can be deformed in the forward time step, a semi-Lagrange algorithm can be used that starts from the last time step and back tracks along the flow lines to update the state.

$$\frac{DT(X, t)}{Dt} = \frac{\partial T(x, t)}{\partial t} + \vec{V} \cdot \nabla T \quad (\text{A.2})$$

The initial temperature in thermal analysis is usually set to the ambient temperature. A convection boundary condition generates a boundary flux q [w/m^2] on all external surfaces. This flux is computed from Eq. A.3 with convection coefficient h . Examples of equations that could be used to evaluate the convection coefficient h , are given in Eq. A.4 [56] for steels and A.5 [57] for aluminum in [$w/m^2 \circ K$]. T is in [$\circ K$].

$$q = h_c(T - T_{\text{ambient}}) \quad (\text{A.3})$$

$$h_c = 5 + 0.05(T - 300) + 6 \times 10^{-7}(T - 300)^3 \quad (\text{A.4})$$

$$h_c = 7.2 - \frac{355000}{T^2} + 0.001 T \quad (\text{A.5})$$

A.1.1 Heat Source Models

For welding, heat source models try to formulate the boundary conditions and/or heat sources near the weld pool to take either transient flux or temperature. The double ellipsoid, added-filler-metal and Ohji heat source models which are used in this thesis, are discussed below.

Double Ellipsoid Model

The double ellipsoid model [58] is the heat source model used most frequently in published papers. This model prescribes a distribution of power density in a weld pool shape characterized by four parameters: width, depth, front length, and rear length of the double ellipsoid as shown in Fig. A.1. Heat power density used in the each of front and rear sections of the double ellipsoid is defined by Eq.s A.6 and A.7 respectively.

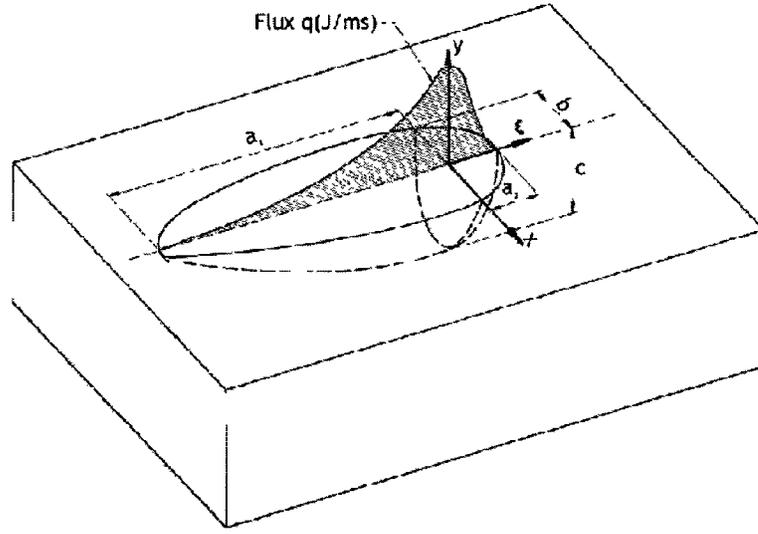


Figure A.1: Double ellipsoid model showing semi-axes length parameters.

$$q(x, y, \xi) = \frac{f_f 6\sqrt{3} Q}{abc \pi \sqrt{\pi}} e^{-3x^2/b} e^{-3y^2/c} e^{-3\xi^2/a_2} \quad (\text{A.6})$$

$$q(x, y, \xi) = \frac{f_r 6\sqrt{3} Q}{abc \pi \sqrt{\pi}} e^{-3x^2/b} e^{-3y^2/c} e^{-3\xi^2/a_1} \quad (\text{A.7})$$

Q is heat input rate (J/s), i.e., $Q = \mu VI$ where μ is efficiency, V is voltage, and I is current of welding arc, (a_1, a_2, b, c) are double ellipsoid parameters, f_f is the fraction of Q in front and f_r is the fraction of Q in rear with $f_f + f_r = 2$. Values of $f_f = 0.6$ and $f_r = 1.4$ were found to provide the best agreement between the measured and calculated values [58]. A double ellipsoid power density distribution with $(0, 0, 0)$ at the weld pool center and semi-axis (a_1, a_2, b, c) parallel to (z, z, x, y) , gives the capability to adjust arbitrarily the double ellipsoid parameters to match a known weld pool shape. In a moving coordinate system (x, y, ξ) in which ξ is the moving in the z direction and v is welding speed [m/s], ξ can be substituted using Eq. A.8 for

a moving ellipsoid in the fixed coordinate system (x, y, z) .

$$\xi = z + v\Delta t \quad (\text{A.8})$$

Added-Filler-Metal Model

This model basically prescribes temperature to specify the heat input. The code defines active nodes on the boundary of the filler metal and part, and prescribes the temperature to these nodes. Three temperatures: T_{part} , $T_{previous}$ and $T_{interior}$ are used to model the distribution of power in the filler metal. T_{part} is the temperature at the boundary which is usually $T_{liquidus}$, $T_{previous}$ is the temperature at the back of the weld pool and is higher than $T_{liquidus}$, and $T_{interior}$ is the temperature in the interior of the weld pool that has the highest temperature of the three temperatures. The values $T_{liquidus} = 1700 \text{ }^\circ K$, $T_{previous} = 1800 \text{ }^\circ K$, and $T_{interior} = 1900 \text{ }^\circ K$ could be used for steels. The sum of Lagrange multipliers for these prescribed temperatures are a measure of the weld power. This power should agree with the real weld power, i.e., μVI .

Ohji Model

Ohji [59] [60] [61] developed a quasi-static weld pool model that predicts the weld pool shape by solving the conservation of energy, mass and momentum. The Ohji weld pool model takes into account arc pressure distribution, surface tension of the liquid metal-air interface and the hydrostatic stress for the direction of gravity for weld pool shape prediction. This model uses its own dynamic mesh fixed in space and the material points flow through the Ohji mesh. Because the Ohji mesh is a small region around the arc position compared to the total mesh, the Ohji mesh can be a fine mesh with a high resolution. This model is a quasi-static model and the results can be mapped from a single solve.

A.2 Microstructure of Low-Alloy Steels

Given the transient temperature field, the evolution of microstructure near a weld pool can be solved using algorithms described in [62] and [63]. These algorithms extend the theory of Kirkaldy in [64]. The essential idea is a pseudo-Iron-Carbon phase diagram in which all lines and points are functions of the alloy composition used to estimate the equilibrium microstructure of the alloy at any time and temperature. The kinetics of the transformation of the austenite or gamma phase to ferrite, pearlite and/or bainite phases that tries to drive the microstructure toward an equilibrium state is governed by ordinary differential equations (ODE) such as Eq. A.9. The transformation of austenite or gamma phase to martensite is governed by the Koistinen-Marburger [65] equation which is an algebraic equation shown in Eq. A.10.

$$\begin{aligned} \frac{dX}{dt} &= \frac{2^{(G-1)/(2)} \Delta_{Fe} T^3 \exp(-23,000/RT) X^{2(1-X)/3} (1-X)^{2X/3}}{COM} \\ COM &= 59.6 \%Mn + 1.45 \%Ni + 67.7 \%Cr + 244 \%Mo \\ \Delta_{Fe} T &= A_3 - T \quad T < A_3 \end{aligned} \quad (A.9)$$

X : Fraction of equilibrium ferrite, not the fraction of volume that is ferrite.

G : ASTM grain size index, $N = 2^{G-1}$ where N is the number of grains in $0.01 [in^2]$.

The evolution of ferrite stops when the equilibrium volume fraction is reached. Also special starting procedures are required to start the integration of this ODE from $X = 0$.

$$V_{martensite}(X, t) = 1 - \exp[-b(M_s - T_q(X, t))] \quad (\text{A.10})$$

M_s : Martensite start temperature.

$T_q(X, t)$: Quench temperature at time t and position X .

b : 1.1×10^{-2} units [$^{\circ}K^{-1}$].

Growth of austenite grain size g in meters is assumed to obey the ODE A.11 given by Ashby and Easterling in [66].

$$\frac{dg}{dt} = \frac{1}{2g} k \left[\exp \left(-\frac{Q}{RT(t)} \right) \right] \quad (\text{A.11})$$

where t is time, k is a constant that must be evaluated from experiment, Q is the activation energy for self-diffusion in austenite and T is temperature in $^{\circ}K$. If carbonytrides are present that pin austenite grain boundaries, austenite grain growth does not start until the carbo-nitrides have dissolved. On cooling austenite grain growth is assumed to halt as soon as ferrite begins to precipitate. The argument for this is that the ferrite pins the austenite grain boundaries. The austenite grain size growth in a thermal cycle is described in detail in [63].

A.2.1 Hardness Evaluation for Low-Alloy Steels

The Vickers hardness of the alloy is H computed by Eq. A.12 where X_m , X_b and X_{fp} are the volume fractions of martensite, bainite and ferrite-pearlite. The Vickers hardness of martensite, H_m , bainite, H_b and ferrite-pearlite, H_{fp} are computed by Eq. A.13, A.14, and A.15 taken from [63]. V designates cooling rate at 1000 [$^{\circ}K$] in

[°K/s] and log is logarithm to base 10.

$$H = H_m X_m + H_b X_b + H_{fp} X_{fp} \quad (\text{A.12})$$

$$\begin{aligned} H_m = & 127.0 + 949.0 * C + 27.0 * Si + 11.0 * Mn + \\ & 8.0 * Ni + 16.0 * Cr + 21.0 * \log V \end{aligned} \quad (\text{A.13})$$

$$\begin{aligned} H_b = & -323.0 + 185.0 * C + 330.0 * Si + 153.0 * Mn + \\ & 65.0 * Ni + 144.0 * Cr + 191.0 * Mo + [89.0 + 53.0 * C - \\ & 55.0 * Si - 22.0 * Mn - 10.0 * Ni - 20.0 * Cr - \\ & 33.0 * Mo] * \log V \end{aligned} \quad (\text{A.14})$$

$$\begin{aligned} H_{fp} = & 42.0 + 223.0 * C + 53.0 * Si + 30.0 * Mn + \\ & 12.6 * Ni + 7.0 * Cr + 19.0 * Mo + [10.0 + \\ & 19.0 * C - 55.0 * Si + 4.0 * Ni + 8.0 * Cr + \\ & 130.0 * V] * \log V \end{aligned} \quad (\text{A.15})$$

A.3 Stress Analysis in Welding

The stress analysis is quasi-static because inertial or dynamic forces are sufficiently small that they can be neglected. Therefore at each instant of time, the domain is in static equilibrium. However, the temperature is time dependent and therefore the thermal strain due to thermal expansion is time dependent. If microstructure evolution is included in the model, then microstructure is time dependent. When a

phase changes, then the specific volume, V , or density, ρ , of the phase changes. The incremental volumetric strain tensor in a time step is $\mathbf{I}\Delta V/V$.

In a time step, the multiplicative deformation gradient F can be written as $F = F_{el}F_{pl}F_{\Delta V}$. The deformation gradient due thermal expansion and phase changes is $F_{\Delta V} = \mathbf{I} + \mathbf{I}\Delta V/V$. The deformation gradient due to visco-plastic deformation is F_{pl} . The deformation gradient due to elastic deformation is $F_{el} = FF_{\Delta V}^{-1}F_{pl}^{-1}$. The reference state or initial condition for each time step is the state at the start of the time step which is constrained to be an equilibrium state.

Given the density ρ , and the fourth order elasticity tensor D_{el} as a 6×6 matrix, the body force b and the Green-Lagrange elastic strain $\varepsilon_{el} = (F_{el}^T F_{el} - \mathbf{I})/2$, VrWeld solves the conservation of momentum equation of the end of each time step that can be written in the form of Eq. A.16 in which inertial forces, $\rho \ddot{x}$, are ignored.

$$\begin{aligned}\nabla \cdot \sigma + b &= 0 \\ \sigma &= D_{el}\varepsilon_{el} \\ \varepsilon_{el} &= (F_{el}^T F_{el} - \mathbf{I})/2\end{aligned}\tag{A.16}$$

VrWeld solves the associated system of non-linear partial differential equations for a visco-thermo-elasto-plastic stress-strain relationship for each time step using theory and algorithms developed by J. C. Simo and his colleagues. For details see [67]. The initial state for each time step is assumed to be at equilibrium.

Appendix B

Optimization Methods

B.1 Derivative-Based Optimization

The main task of optimization is to compute the derivative of an objective function (ϕ) wrt control vector (u), i.e., $d\phi/du$. Taking account of the dependence of state function (Q), differentiating ϕ gives Eq. B.1.

$$d\phi = \phi_Q dQ + \phi_u du \implies \frac{d\phi}{du} = \phi_Q \frac{dQ}{du} + \phi_u \quad (\text{B.1})$$

The objective function, ϕ , is defined by the user and the terms ϕ_Q and ϕ_u are explicit derivatives of that objective function wrt Q and u . The objective functions are usually defined independent from control parameters and therefore the term ϕ_u is usually zero. The only term that needs to be determined is dQ/du [12].

Substitute for dQ by differentiating the state equation, $dQ = F_Q dQ + F_u du$, giving Eq. B.2

$$(I - F_Q) \frac{dQ}{du} = F_u \quad (\text{B.2})$$

Hence we have the problem given in Eq. B.3

$$\begin{aligned} & \text{find } \phi_Q \frac{dQ}{du} \\ \text{such that } & (I - F_Q) \frac{dQ}{du} = F_u \end{aligned} \quad (\text{B.3})$$

There are many approaches to solve or approximate this system of equation. These are discussed in the following section.

B.1.1 Finite Difference

The simplest strategy is to use finite difference machinery by changing u and observing the behavior of the Q . We might use multi-projects to generate some directional derivative in a certain direction of the control space. We could then fit a polynomial to those solutions to make an approximate response surface of the objective function. This is a very simple and robust method to approximate the derivative and most of this machinery is used for other optimization approaches or validation of other derivative-based methods. The disadvantage is that this algorithm gives no information on the gradient of the objective function wrt to design parameters and might need a great number of tests to approximate the derivative. This is slow and CPU costs are high for a given accuracy in problems including many design parameters.

B.1.2 Direct Differentiation

This method is to form an explicit derivative of the state variables from the discretized constitutive equation wrt the control parameter [6]. For example, the conservative of energy equation is discretized as Eq. B.4 assuming there is no heat sink/source and

no C_p .

$$\begin{aligned} \int_V B^T \kappa B \bar{T} dv + \int_A R^T \bar{q} dA &= 0 \\ \sum_{GV_e} B^T \kappa B \bar{T} W J + \sum_{GA_e} R^T \bar{q} w j &= 0 \end{aligned} \quad (\text{B.4})$$

Where GV_e and GA_e denotes Gauss points in the element volume and element surface. B and R are the interpolation matrix for gradient and value of nodal quantities. W and w are the Gaussian weighs for element and surface Gauss points. J and j are the Jacobian for element and surface. κ is the heat conduction coefficient matrix. q is the surface flux and T is the nodal temperature, i.e.,state variables.

We are looking for the derivative of the state equation wrt control parameter u and the direct differentiation method takes the derivative of Eq. B.4 wrt u as given in Eq. B.5.

$$\begin{aligned} \sum_{GV_e} \left\{ \left[\frac{dB^T}{du} \kappa B + B^T \frac{d\kappa}{du} B + B^T \kappa \frac{dB}{du} \right] \bar{T} W J + [B^T \kappa B \bar{T} W] \frac{dJ}{du} \right\} \\ + \sum_{GA_e} \left\{ R^T \frac{d\bar{q}}{du} w j + R^T \bar{q} w \frac{dj}{du} \right\} = 0 \end{aligned} \quad (\text{B.5})$$

For simplicity, let's assume that the u affects the flux only and hence all terms in Eq. B.5 become zero except for $R^T \frac{d\bar{q}}{du} w j$. The flux q , for instance, comes from a differentiable double ellipsoid function [23]. Substituting in Eq. B.1 gives $d\phi/du$ in an explicit form for evaluation.

Optimization using software with direct differentiation algorithms is expected to be much faster and more accurate because it provides the gradient of the objective function. However, the disadvantage is that it requires the code to be written to

compute the gradient of the state equation or residual wrt to control parameters. If one can provide such machinery, this method is a good choice. This method therefore might not be practical for a complicated problem.

B.1.3 Adjoint Methods

There exist many direct methods for solving the system of equations in Eq. B.3. The adjoint method is an alternative way of computing the gradient of ϕ wrt u . Generally, $\phi(x)$ maps $x \in R^m$ to R^p . The adjoint method saves much work when m is large and p is small. Otherwise the direct methods work much better. In most practical cases, ϕ is a scalar-valued function that means $p=1$ and therefore the adjoint method is useful [12]. Two viewpoints are presented in literature for the adjoint method; duality and Lagrange multiplier viewpoint [68].

Duality Viewpoint

let's reform Eq. B.3 as given in Eq. B.6 for a governing equation $KQ = R$.

$$K \frac{dQ}{du} = \frac{dR}{du} \tag{B.6}$$

To switch to the adjoint form, multiply both sides of the equation B.6 by $(d\phi/dR)^T$;

$$\frac{d\phi^T}{dR} K \frac{dQ}{du} = \frac{d\phi^T}{dR} \frac{dR}{du} \tag{B.7}$$

Simplifying this equation;

$$\frac{d\varphi}{dR}{}^T K \frac{dQ}{du} = \frac{d\varphi}{du} \quad (\text{B.8})$$

The only way this equation remains true is if $\frac{d\varphi}{dR}{}^T K$ is equal to $\frac{d\varphi}{dQ}{}^T$. Therefore following system of equations should be satisfied;

$$\frac{d\varphi}{dR}{}^T K = \frac{d\varphi}{dQ}{}^T \quad (\text{B.9})$$

Taking the transpose of both sides, it becomes;

$$K^T \frac{d\varphi}{dR} = \frac{d\varphi}{dQ} \quad (\text{B.10})$$

$d\varphi/dQ$ is the explicit derivative of the objective function and this system of equations, therefore, is solved for $d\varphi/dR$. The work of this system is proportional to p and a remarkable amount of work is saved if p is much smaller than m comparing to the direct method. $p=1$ is very often the case in real problems.

Once the result for $d\varphi/dR$ is computed from Eq. B.10, the chain rule gives $d\varphi/du$ by following relation;

$$\frac{d\varphi}{du} = \frac{d\varphi}{dR}{}^T \frac{dR}{du} + \varphi_u \quad (\text{B.11})$$

dR/du can be evaluated with an automatic differentiation algorithm that evaluates the change in R wrt control parameter u and there is no solve involved.

In effect, the adjoint method repeats evaluation of dR/du when u is perturbed in compare to the direct method that repeats the linear equation solving for dQ/du when u is perturbed [68].

Lagrange Multiplier Viewpoint

Lagrange multipliers, by definition, is a method to maximize/minimize a function subject to constraints. If the constraint is the state equation $KQ=R$, the Lagrange function is written in Eq. B.12 where $\lambda_{n \times p}$ is Lagrange multiplier [18].

$$L(Q, u) = \varphi(Q, u) - \lambda^T(KQ - R) \quad (\text{B.12})$$

$KQ-R$ is the residual of the state equation and let's, therefore, call it Res .

$$L(Q, u) = \varphi(Q, u) - \lambda^T(Res) \quad (\text{B.13})$$

The chain rule for general perturbation in Q and u ;

$$dL = \left(\frac{\partial \varphi}{\partial Q} - \lambda^T \frac{\partial Res}{\partial Q} \right) dQ + \left(\frac{\partial \varphi}{\partial u} - \lambda^T \frac{\partial Res}{\partial u} \right) du \quad (\text{B.14})$$

If λ^T is chosen to satisfy the Adjoint equation;

$$\left(\frac{\partial \varphi}{\partial Q} - \lambda^T \frac{\partial Res}{\partial Q} \right) dQ = 0 \implies \left(\frac{\partial Res}{\partial Q} \right)^T \lambda = \left(\frac{\partial \varphi}{\partial Q} \right) \quad (\text{B.15})$$

then;

$$dL = \left(\frac{\partial \varphi}{\partial u} - \lambda^T \frac{\partial Res}{\partial u} \right) du \implies \frac{dL}{du} = \left(\frac{\partial \varphi}{\partial u} - \lambda^T \frac{\partial Res}{\partial u} \right) \quad (\text{B.16})$$

The main drawbacks for the adjoint method arise in implementation of a control problem. If we have a chain of state vectors $Q_0, Q_1, Q_2, Q_3, \dots, Q_n$ in the control problem, the forward states are calculated in order as in an ordinary simulation and

each state must be stored. Then, the adjoint states $R_1, R_2, R_3, \dots, R_n$ are calculated, proceeding backwards through the sequence. Two main drawbacks are that we have to compute the forward states because the adjoint calculation depends on them, and we have to store the results which could result in a huge memory requirement. In addition, a code to step backward in time must be implemented [14].

B.1.4 Derivative-Based Optimization Drawbacks

There exist a great number of derivative-based optimization algorithms. The algorithms using the derivative are somewhat more powerful than those using only the function values [17]. However they are limited in that firstly both the objective function and the design parameters have to be continuous and secondly if the objective function contains multiple extrema, then the gradient methods converge to the nearest local extremum with little chance to find the global extremum. In these cases, derivative-free approaches can be used to solve the problem [18]. This opens the discussion on derivative-free methods for a general platform optimizer and this thesis focuses on these methods.

B.2 Derivative-Free Optimization

Derivative-free optimization methods can be categorized into (a) local-minimum search methods using an accurate model, (b) global-chance-minimum search methods using an accurate model (d) local-minimum search methods using an approximation model, (e) global-chance-minimum search methods using an approximation model. There are tens of algorithms in each category and some of them are discussed below.

B.2.1 Nelder-Mead Downhill Simplex

Nelder and Mead (N-M) introduced the downhill simplex method (1965), which does not require the calculation of derivatives. A simplex is simple geometry such as a triangle for a 2D problem and the downhill simplex method works based on the simplex. Only one point of the simplex, P_0 , is specified by the user as a start point. The other points are found by machine to form a simplex around the user point. The goal of this method is to move the simplex until it surrounds the minimum, and then to contract the simplex around the minimum until it is within an acceptable tolerance. The main step used to find the local minimum is based on the reflection. Let us take the simplex ABC in Fig .B.1 [69] as an initial simplex. Evaluation of result shows point A as the lowest value among (A, B, C). The reflection of A is a line connecting A to the midpoint of BC and continues by the same length to D.

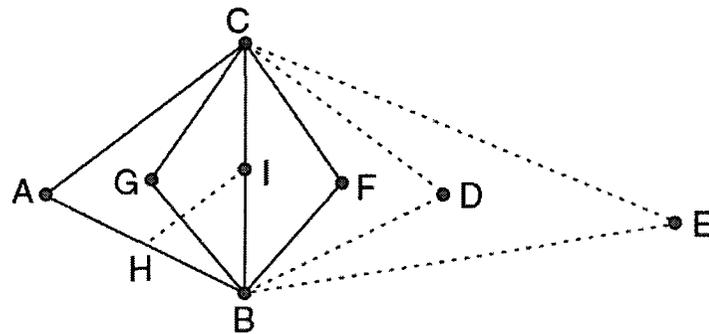


Figure B.1: Manipulation of the basic 2D simplex to find the minimum by N-M downhill simplex method.

If evaluation of D shows lower value, we are in a right direction and next point would be E (expansion) otherwise the new simplex becomes BCD and reflection repeats for the lowest point in the new simplex BCD.

If evaluation of D is close to A then two new points F and G are found and the smallest F or G is kept to form a new simplex (contraction).

If neither F nor G is smaller than A , AC moves to the B which is the next lowest of ABC and forms HIB as the new simplex (shrinkage).

Fig. B.2 shows an example solved by N-M algorithm. This algorithm is not fast, but it has a certain robustness that makes it attractive. The N-M algorithm may get stuck in local minimum and it is better to be combined with some global search algorithm to find the global minimum [69].

A variety of improved and automated N-M algorithms have been developed [70]. The basic assumption for all such algorithms is getting one result at a time for decision making. Having an optimizer platform using multi-core computers can benefit from a version of a parametric-design algorithm to make it efficient and fast. The idea of parametric-design DOE optimization is discussed further.

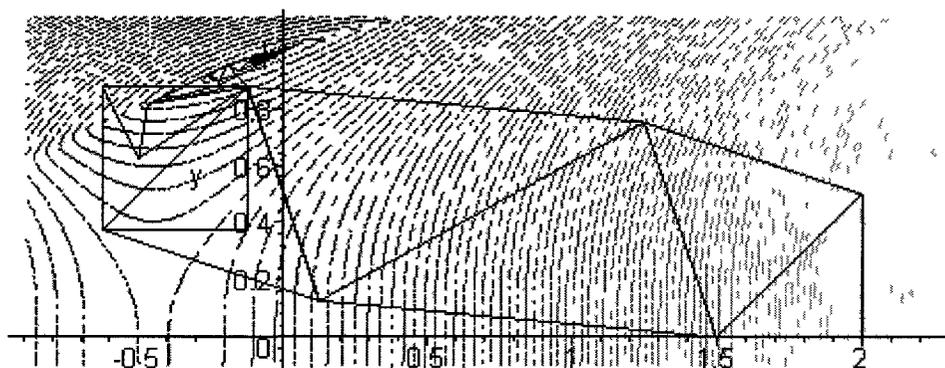


Figure B.2: An example solved by N-M algorithm.

B.2.2 Direct-Search Algorithms

In the 1990s, direct-search methods were presented in the context of parallel computing [33]. This introduced re-design of such algorithms for multi-cores platforms. Kolda, Lewis and Torczon have proposed a basic algorithm called “*Compass search*” and an advanced one called “*GSS*” for direct-search optimization in [35]. Their ideas are summarized in the pseudo-code below;

Given an objective function, a DOE matrix formed for a starting point with a trial step size, and a supporting framework:

1. Run DOE matrix and assess results.
2. If there is a new minimum with a sufficient decrease in the results, then
 - Move to the minimum as a new trial point and go to 1.
3. If there is no new minimum point in the results, then
 - Test for convergence.
 - If no convergence, refine the trial step size and go to 1.
 - If convergence, write report and stop.

Fig. B.3 illustrates steps taken by the “*Compass search*” algorithm using 4 cores for evaluation of results [35].

B.2.3 Least-Square Algorithm

Least squares methods, generally, approximate over-determined systems of equations such that the sum of the squares of the residuals becomes minimum. This method is the basis for the regression models that are frequently used to approximate the behavior of a system as a mathematical function of independent variables. The idea of using a least-square approximation in direct-search optimization, is presented as the following pseudo-code:

Given an objective function, starting point, parameter step-sizes and parameter bounds construct a mesh or grid to discretize the design space:

1. Run DOE matrix and assess results.
2. Construct the least-square regression model for the objective function on the assessed results.

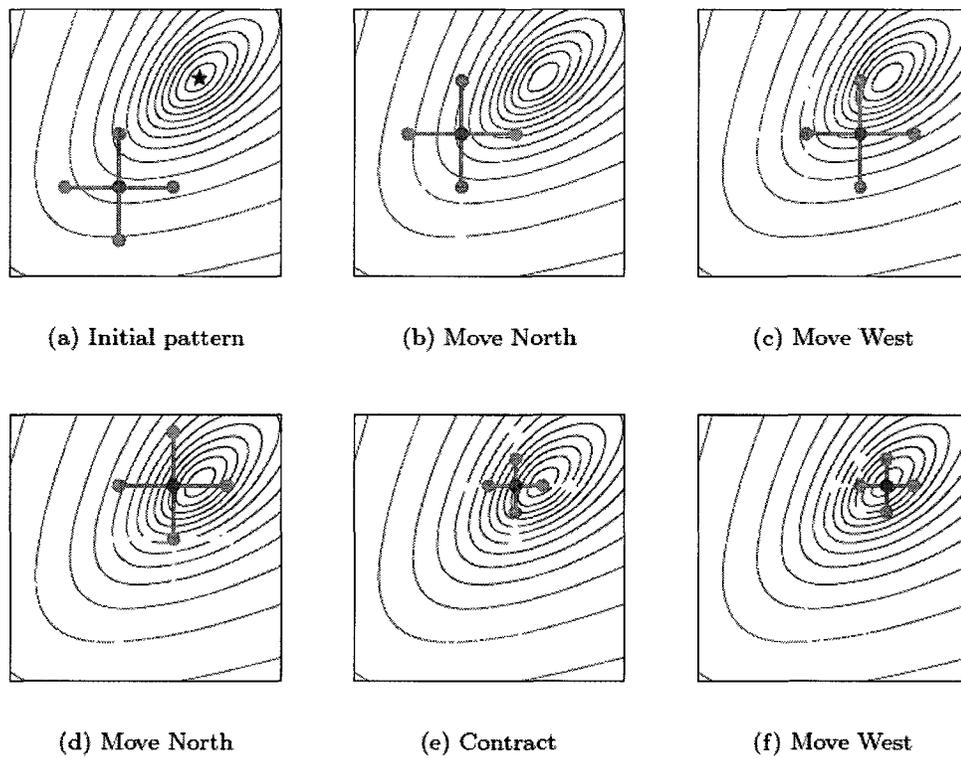


Figure B.3: Steps taken by the *Compass search* algorithm using 4 cores for evaluation of results.

3. Test for convergence, then

If converged, write report and stop.

If not, find values of parameters for the minimum of the regression equation.

If the minimum from regression approximation is smaller than accurate results in the DOE matrix then run the configuration of the regression-found minimum using the accurate model.

If the accurate model and the regression model agree within a certain tolerance, move to the new configuration.

If not, move to the minimum in the regression-tested region. Find the closest point on the mesh (or 2R-refined mesh ¹) and perturb based on the gradient computed from the least-square approximation on the node picked.

4. Go to 1

B.2.4 Line Search Algorithms

There exist a variety of line-search algorithms that begin at some trial point, choose a direction to move, then move in that direction until the objective function begins to increase. Next the procedure is repeated in another direction. Finding a sensible direction to move is critical to algorithm convergence and has spawned a variety of approaches.

A simple approach to line minimization is the coordinate search method (Schwefel, 1995). It starts at an trial point on the objective surface, then does a line minimization along the axis of one of the variables. Next, it selects another variable and does

¹One point is added in the middle of the existing mesh line that connects two points on the current mesh.

another line minimization along that axis. This process continues until a line minimization is done along each of the variables. Then the algorithm cycles through the variables until an acceptable solution is found. Fig. B.4 shows a possible path the algorithm might take in a quadratic objective surface. In general, this method is a slow method [69].

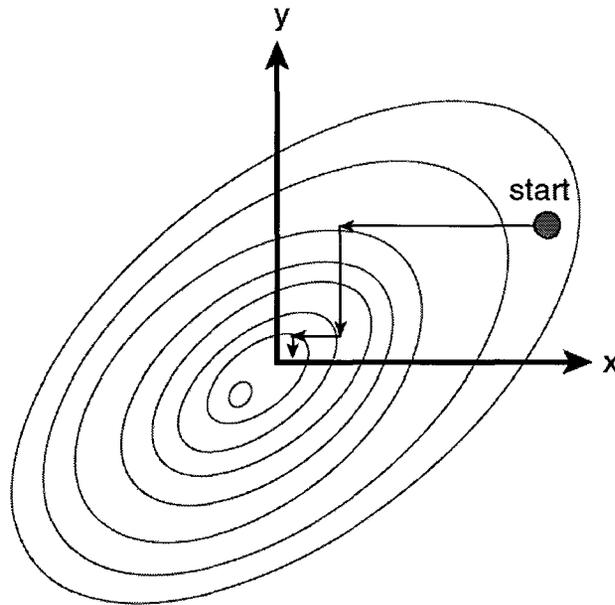


Figure B.4: A possible path the algorithm might take in a quadratic objective surface by the coordinate search method.

Rosenbrock (1960) developed a method that does not limit search directions to be parallel to the variable axes. The first iteration of the Rosenbrock method uses coordinate search along each variable to find the first improved point (see Fig. B.5). The coordinate axes are then rotated until the first new coordinate axis points from the starting location to the first point. Gram-Schmidt orthogonalization finds the directions of the other new coordinate axes based on the first new coordinate axis. A coordinate search is then performed along each new coordinate axis. As before, this process iterates until an acceptable solution is found [69].

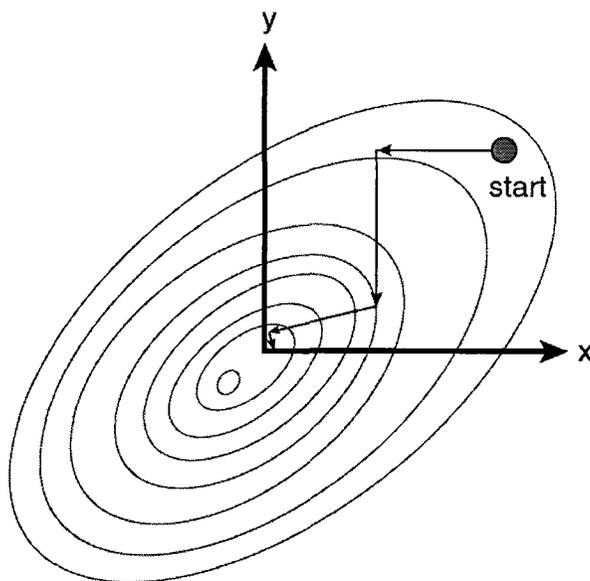


Figure B.5: A possible path the algorithm might take in a quadratic objective surface by the variable coordinate axes search method.

The steepest descent algorithm originated with Cauchy in 1847. It starts at an trial point on the objective surface and minimizes along the direction of the gradient. By definition, the new gradient formed at each iteration is orthogonal to the previous gradient and therefore the path is orthogonal to the contours and any path is orthogonal to the previous and next path. The concern about this method is that if the valley is narrow, this algorithm could bounce from side to side for many iterations before reaching the bottom. Fig. B.6 shows a possible path of the steepest descent algorithm [69].

Powell (1964) [69] developed a method that finds a set of line minimization directions that are linearly independent, mutually conjugate directions. The direction ensuring the current direction does not spoil the minimization of the prior direction is the conjugate direction. The conjugate directions are chosen so that the change in the gradient of the objective function remains perpendicular to the previous direction. Fig. B.7 shows a possible path of the conjugate gradient algorithm.

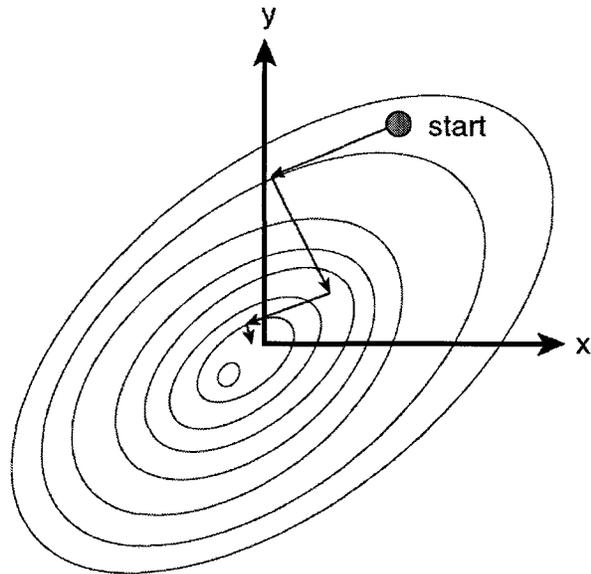


Figure B.6: A possible path the algorithm might take in a quadratic objective surface by the steepest descent search method.

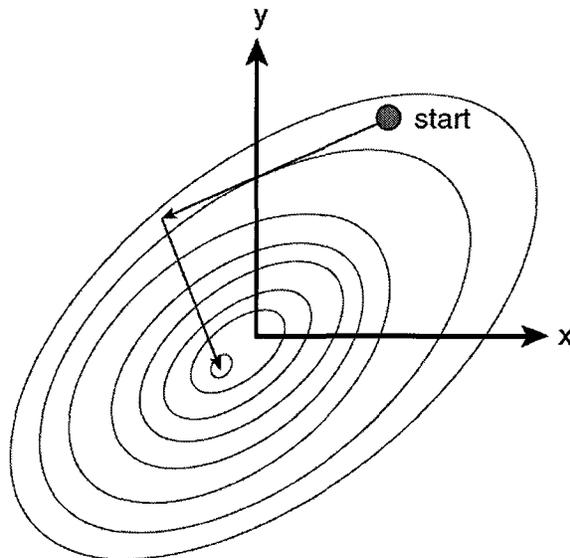


Figure B.7: A possible path the algorithm might take in a quadratic objective surface by the conjugate gradient search method.

Many other search algorithms developed/improved in this fashion such as Bundle methods, the Ellipsoid method, the FrankWolfe method, interior point methods, quasi-Newton methods, and so on. For all of them, the objective function should be convex over the region of interest, then any local minimum will also be a global minimum [34]. When the problem is not convex, all of available algorithms presented are based on the random search to provide a chance of capturing the global minimum. Differences in the global search algorithms come from how the algorithm uses the randomness and the number of random point tested to make the search algorithm as efficient as possible for the problem of interest. Again there is no one-best algorithm and the method selection depends on the problem characteristics. Current random search algorithms available in the literatures could be named as Ant-colony optimization, Bat algorithm, Beam search, Bees algorithm, Cuckoo search, Dynamic relaxation, Evolution strategy, Filled function method, Firefly algorithm, Genetic algorithms, Harmony search, Hill climbing, IOSO, MonteCarlo, Particle swarm optimization, Quantum annealing, Simulated annealing, Stochastic tunneling, and Tabu search [34].

B.2.5 Surrogate/Meta Model Algorithms

Optimization requires evaluation of a design objective and the constraints as a function of design variables. All of the algorithms discussed so far work on the sample space to use the smallest possible sample space and gain the most information out of that to move toward the minimum. The main limitation of using a large sample space or to test all nodes in the space of parameters is that the evaluation of the results takes time. In effect, the fastest evaluation can test a larger sample space and make a better choice. This introduces another collection of optimization approaches that construct an approximate model such as a surrogate/meta model that mimics the behavior of the accurate model as closely as possible while being computationally

cheap(er) to evaluate [51].

Reference [7] developed a surrogate model based on observing how the function typically behaves on a certain nodes and when predicting for other nodes, they are essentially computing the function value that is most consistent with this estimated typical behavior.

Such a surrogate-based optimization typically includes the following steps;

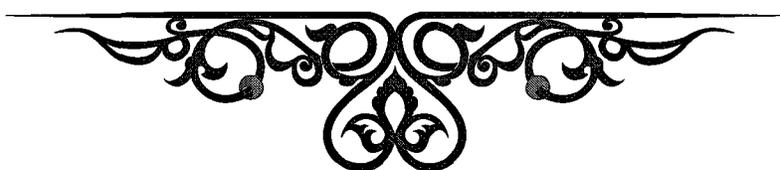
1. Design of experiments (DOE).
2. Evaluate the DOE with an accurate model.
3. Use this evaluation to train a surrogate model to approximate the desired relationships.
4. Evaluate all combinations using the trained surrogate model.
5. Select the optimum result from the surrogate and evaluate it using the accurate model.
6. If the result is not good enough, add it to the existing DOE results and re-train the surrogate model.

The End



حريم عشق را در كه بسي بالائز عقل است
كسي آن آسان بوسد كه جان در آستين دارد

حافظ



The author's interpretation:

*When it comes to love, intellect has no hope.
Alive creatures are the only kind who understand it.*

*Hafez, a Persian lyric poem
(1325 – 1389 CE, Iran)*