Chapter 7

Numerous Design Variables

In previous chapters, the use of simplified models and central finite differencing for the determination of gradient information, when optimising the off-road vehicle's suspension characteristics for ride comfort and handling, was shown to be beneficial. The problems considered, however, looked at only a few multiplication factors to define the suspension characteristics. In this chapter the suspension characteristics are defined by up to 14 design variables, dramatically increasing the complexity of the optimisation problem. The design variables are used to define the non-linear spring and damper characteristics, with these characteristics being optimised for the vehicle's handling and ride comfort. This chapter highlights the complications involved with the higher number of design variables. Poor scaling and sensitivity effects are illustrated in typical optimisation convergence histories, and solutions highlighted. The improved scaling discussed, dramatically helps to improve the convergence history with respect to noise. This chapter thus aims to give the optimisation engineer techniques for identifying and correcting complications associated with gradient-based vehicle suspension optimisation. It is normally these complications that lead to the adoption of less efficient stochastic based optimisation methods. While not all the complications are solved, reasons for the complications are investigated.

Based on the success of the non-linear simplified models describing vehicle handling and ride comfort, to obtain the gradient information for optimisation

problems with two and four design variables, the problem is expanded to 14 design variables. These 14 design variables better describe the shape of the damper characteristics front and rear, and the static gas volume front and rear.

The concept of using the simplified models for the calculation of the gradient information, as proposed by Balabanov and Venter (2004) for finite element structural problems, and Chapter 5 for vehicle suspension optimisation, is now assumed to be sufficiently representative of the full simulation model. This chapter will thus only define the design variables and discuss the optimisation results, with emphasis on the adjustments needed when considering many design variables. The vehicle model used is the same as in Chapter 6, except for the design variables that define the damper characteristics.

7.1 Definition of Design Variables

The front and rear static gas volumes are kept as design variables, defining the non-linear spring stiffness. The design variables that define the damper characteristics are redefined in order to achieve a more accurate description of the required damper characteristics. The standard rear damper characteristic is used, and redefined in terms of piecewise quadratic approximations, as illustrated in Figure 7.1. This gives a very accurate approximation to the measured damper characteristics. The damping force is primarily generated as a result of oil flow through an orifice, and for this reason the quadratic approximation is used to describe the characteristics. The general description of the force generated by oil flow through an orifice can be described by the quadratic relation:

$$F = kv^2 \tag{7.1}$$

where F is the damper force, v the velocity of the relative displacement of the piston (in this case of the suspension strut between the axle and the body), and k a correlation coefficient, dependent on the area and drag factor (C_d) of the orifice.

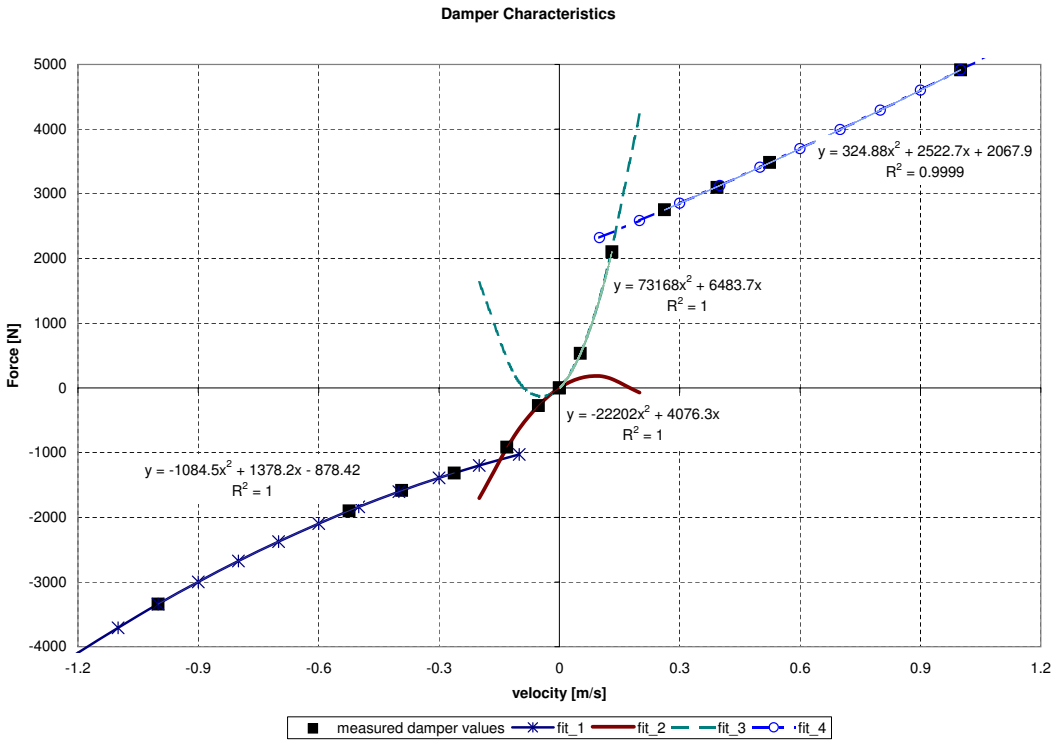


Figure 7.1: Definition of damper characteristics with quadratic approximation to the baseline rear Land-Rover damper

The damper fits with scale factors $sf_{1\rightarrow 6}$ are defined as follows:

$$\begin{aligned} fit_1 &= sf_1(-1084.5v^2 + 1378.2v) - sf_2(878.2) \\ fit_2 &= sf_3(-2220.2v^2 + 4076.3v) \\ fit_3 &= sf_4(7316.8v^2 + 6483.7v) \\ fit_4 &= sf_5(324.88v^2 + 2522.7v) + sf_6(2067.9) \end{aligned} \tag{7.2}$$

The damper force F_{dmp}, using the above piecewise fits to the measured damper characteristic, is defined as follows:

$$\begin{aligned} &if \quad v \leq 0 \\ &F_{dmp} = max(fit_1(v), fit_2(v)) \\ &else \\ &F_{dmp} = min(fit_3(v), fit_4(v)) \\ &end \end{aligned} \tag{7.3}$$

The full damper force velocity characteristic can now be defined in terms of the six scale factors $sf_{1\rightarrow 6}$. These damper scale factors are allowed to range

between 0.1 and 3. The design variables can then be stated as follows:

$$x_{1\to 6} = \frac{sff_{1\to 6}-0.1}{3-0.1}, \qquad x_7 = \frac{gvolf-0.1}{0.6-0.1}, \tag{7.4}$$
$$x_{8\to 13} = \frac{sfr_{1\to 6}-0.1}{3-0.1}, \qquad x_{14} = \frac{gvolr-0.1}{0.6-0.1},$$

with bounds:

$$0.001 \leq x_i \leq 1, \qquad i = 1, ..., 14 \tag{7.5}$$

where sff and sfr denotes the front and rear damper scale factors, $gvolf$ and $gvolr$ the front and rear static gas volumes, of the $4S_4$ suspension system. As before the static gas volumes range between 0.1 and 0.6 liter. All the design variables are then scaled to range from zero and one as suggested by Snyman (2005b).

The normalised objective and constraint functions defined in Chapter 5 are again used for the optimisation of ride comfort and handling.

7.2 Handling Optimisation

The handling optimisation was performed using the middle of the design space as a starting point, the opposite of the 4 design variable optimum (i.e. the infeasible point), and a random point in the design space. The results indicated that design variables 3, 4, 10 and 11 (the scale factors of fit_2 and fit_3, front and rear) all converged to the maximum boundary value of one, while design variables 7 and 14 (the gas volumes) converged to the minimum boundary value of almost zero. However, the other design variables did not change from their initial starting value. Yet when using different values starting values for these variables, different minima $f^*(\mathbf{x})$, less than the above result were obtained. This indicates that design variables 1, 2, 5, 6, 8, 9, 12 and 13 do have an effect on the local minimum found, yet not as strong as 3, 4, 7, 10, 11 and 14. Figure 7.2 indicates the objective function convergence history and the relative summed change in the design variables, and objective function, from one iteration to the next. It could be argued that the convergence/termination criteria are not strict enough allowing

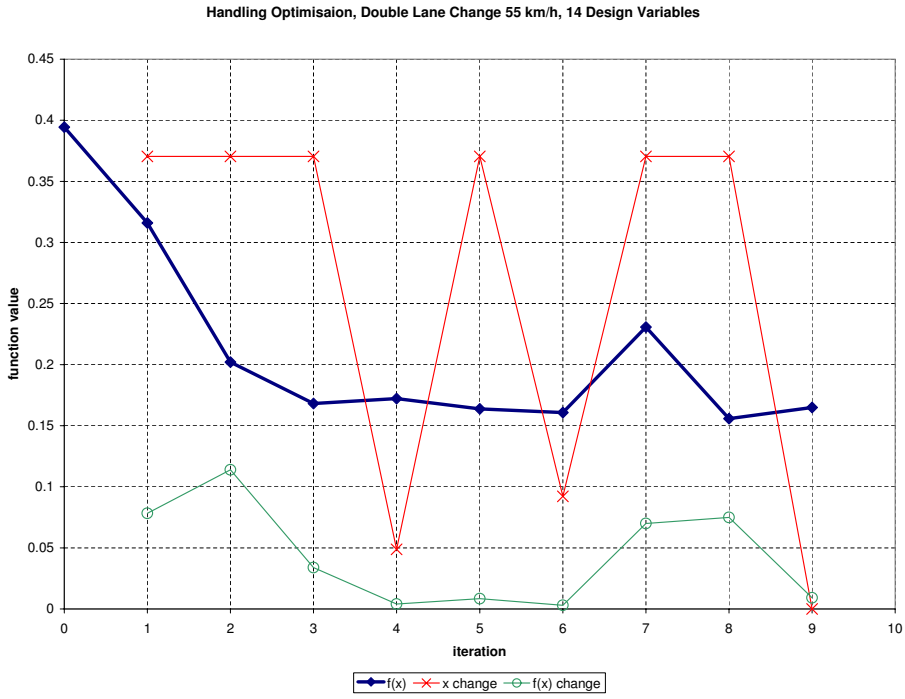


Figure 7.2: Handling 14 design variable optimisation convergence history.

premature termination of the optimisation algorithm on a non-optimum point. The termination criteria were then made 10% of previous, yet the optimisation still converged to the same points. This indicates that the design variable's current scaling flattens out their effect, resulting in an almost zero gradient, or low sensitivity.

This low sensitivity could be overcome by one of two methods. Firstly by rescaling the particular variables that they have the same magnitudes but over a much smaller range, effectively increasing their sensitivity. Alternatively by using a much larger perturbation of the design variables when calculating the gradient by central finite differences. However, Figure 7.3 indicates that the relative change in the design variable is so small, that, a change in the perturbation when calculating the gradient will not work. When observing the change in the normalised objective function values with respect to, for example design variable x_2 (Figure 7.4), it can be seen that there is a definite minimum of the objective function with respect to x_2. The objective function with respect to these design variables, or the design

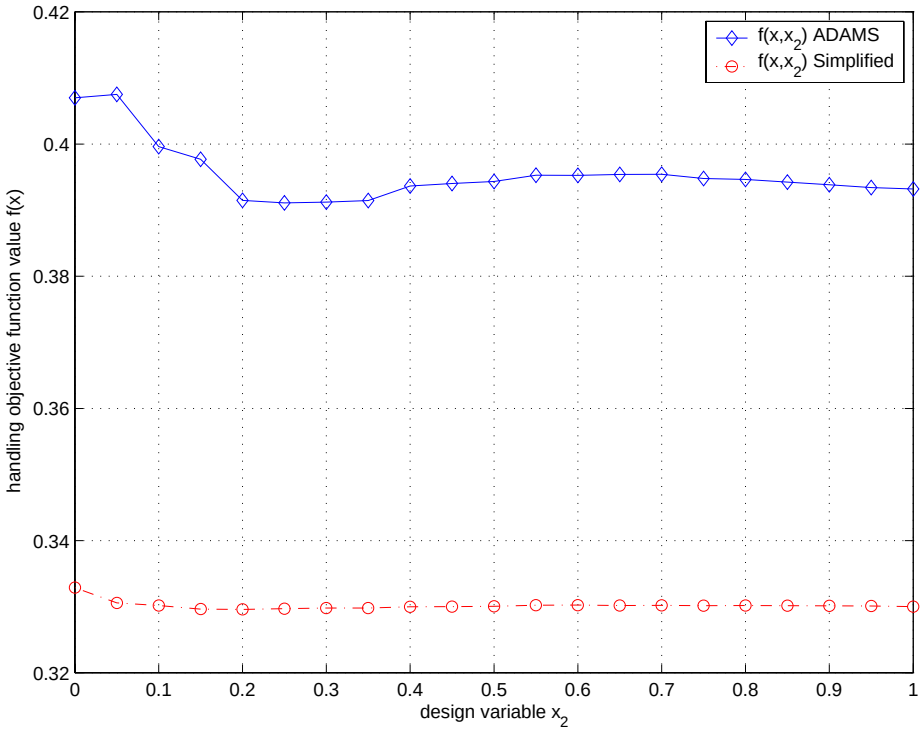


Figure 7.3: Change in objective function value with respect to design variable x_2, when the other design variables are in the middle of the design space

variables themselves has to be rescaled. This highlights that the scaling of the design variables between zero and one, as suggested by Snyman (2005b) does not necessarily guarantee good convergence to a minimum.

The design variables should thus be scaled to have almost equivalent sensitivity, without deviating too far away from similar ranges and magnitudes. This implies that it is more desirable to have an objective function of a spherical nature rather than an elliptic nature. Figure 7.5 illustrates this point, by showing the more direct and faster convergence to the two design variable optimum, when the objective function is scaled to be more spherical, as opposed to the elliptic objective function that has the design variables scaled between zero and one. In the elliptic objective function graph, it can be seen that due to slight errors in the approximate quadratic approximations, the design variables 'jump' around the optimal

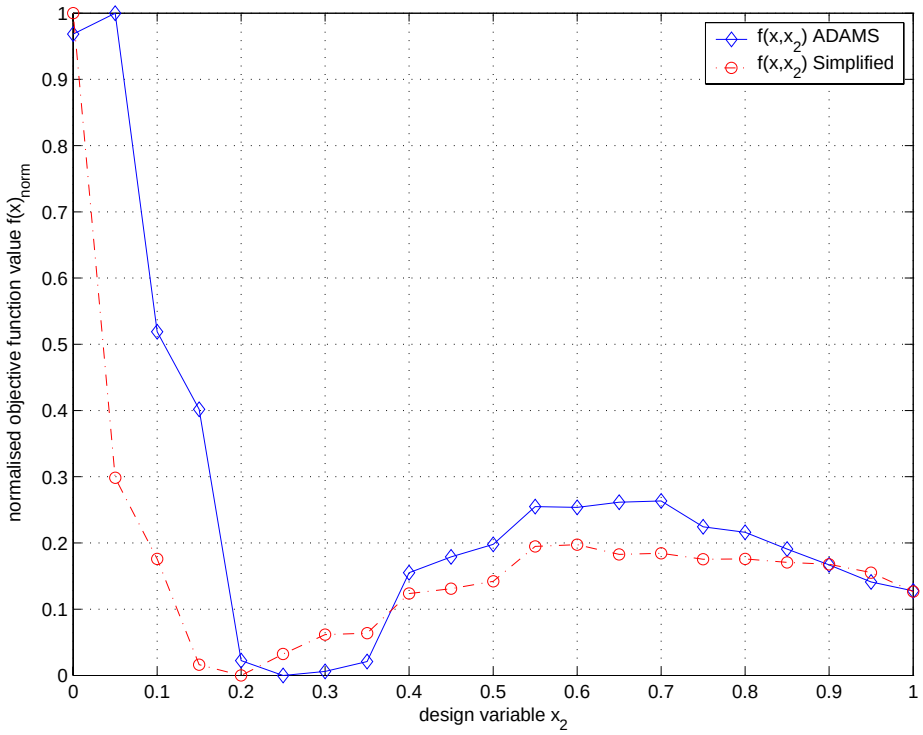


Figure 7.4: Normalised change in objective function value with respect to design variable x_2, when the other design variables are in the middle of the design space

line to the optimum. This jumping is another reason for the spikes seen in the objective function's optimisation convergence history when close to the optimum. Thus, although the objective function's history may appear erratic, the design variables are moving closer to the optimum point in the design space. This effect is amplified when the objective function is severely elliptic, with respect to the design variables, i.e. combination of a steep valley with respect to the one design variable and a shallow valley with respect to the other design variable. Also to consider is that Dynamic-Q constructs successive spherical quadratic approximations to the optimisation problem, thus if the optimisation problem exhibits a more spherical nature, the successive approximations will be a more accurate approximation, resulting in faster convergence to the actual optimum.

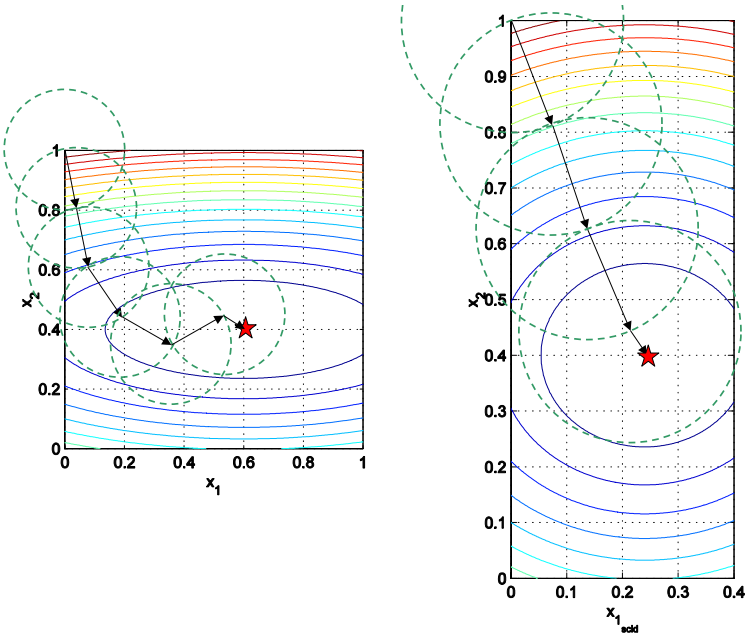


Figure 7.5: Illustration of the effect of ellipticity and sphericality on the convergence to the optimum

7.3 Ride Comfort Optimisation

The optimisation was performed with 14 design variables for ride comfort. Contrary to the handling optimisation case, it was found that design variables 1, 2, 5, 6, 8, 9, 12 and 13 (damper scale factors of fit_1 and fit_4 front and rear) had the largest effect on the erratic behaviour of the objective function value. Figure 7.6 illustrates the convergence history for the first 20 iterations, where it can be seen that changes in design variables 1 and 12, correspond to spikes in the objective function value, while design variable 14 is well behaved. Design variables 1, 2, 5, 6, 8, 9, 12 and 13 all exhibited similar trends that appeared erratic. On closer inspection it was noted that for a small change in the design variable relative to it's allowable range, there is a relatively dramatic change in the objective function value, as observed in Figure 7.6. It is thus proposed that these design variables should be rescaled so that the original normalised range of 0 to 0.4 (resulting in a scale factor range of 0.1 to 1.26) becomes their new 0 to 1 range. This effectively decreases their

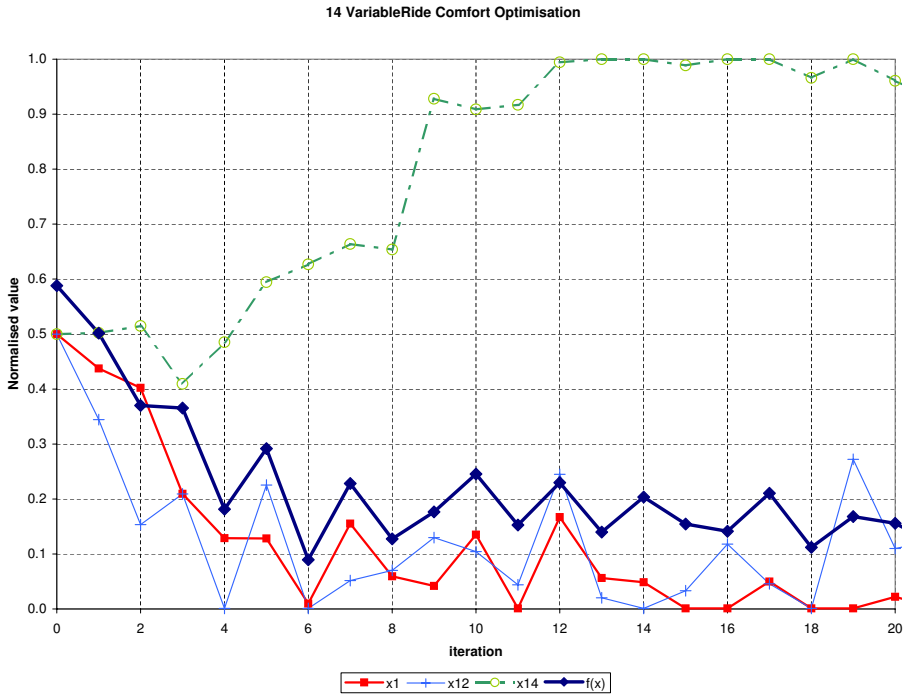


Figure 7.6: Ride comfort optimisation convergence history illustrating first 20 iterations

sensitivity, and flattens out the change in objective function with respect to a change in design variable value. This is opposite to what is needed for the handling optimisation described above. The design variables can then be stated as follows:

$$\begin{aligned} x_{1\to2} &= \tfrac{sff_{1\to2}-0.1}{1.27-0.1}, & x_{3\to4} &= \tfrac{sff_{3\to4}-0.1}{3-0.1}, \\ x_{5\to6} &= \tfrac{sff_{5\to6}-0.1}{1.27-0.1}, & x_{7} &= \tfrac{gvolf-0.1}{0.6-0.1}, \\ x_{8\to9} &= \tfrac{sfr_{1\to2}-0.1}{3-0.1}, & x_{10\to11} &= \tfrac{sff_{3\to4}-0.1}{1.27-0.1}, \\ x_{12\to13} &= \tfrac{sff_{5\to6}-0.1}{1.27-0.1}, & x_{14} &= \tfrac{gvolr-0.1}{0.6-0.1}, \end{aligned} \tag{7.6}$$

with bounds:

$$0.001 \leq x_i \leq 1, \qquad i = 1, ..., 14 \tag{7.7}$$

where sff and sfr denotes the front and rear damper scale factors, and $gvolf$ and $gvolr$ the front and rear static gas volumes of the $4S_4$ suspension system, as defined in Section 7.1. The results obtained in Figure 7.7 for the

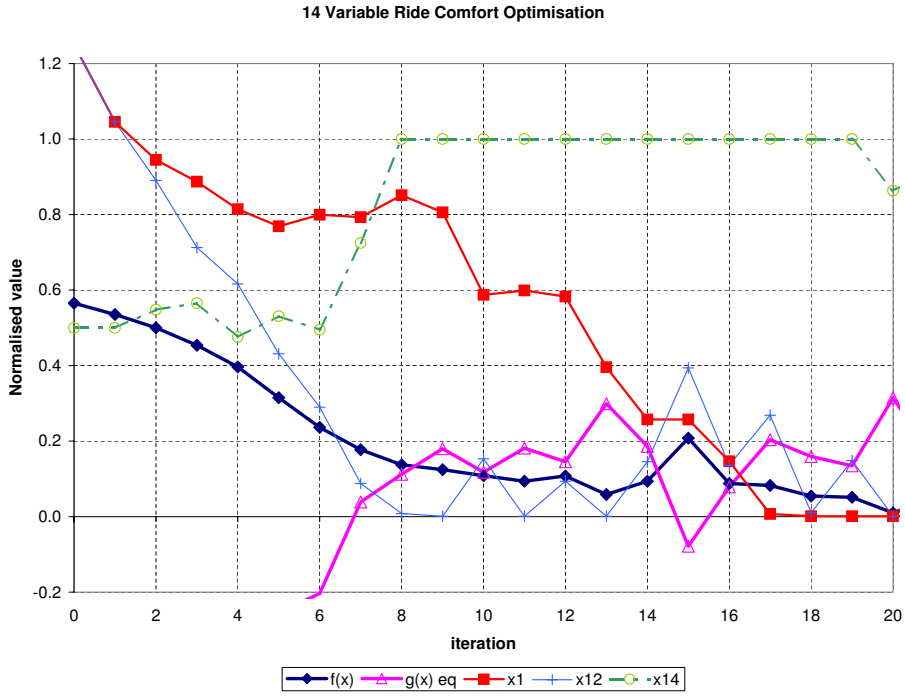


Figure 7.7: Ride comfort optimisation convergence history with rescaled design variables

rescaled problem show a dramatic improvement in the objective function's optimisation convergence history compared to Figure 7.6. However, the inequality constraint is poorly satisfied for most of the optimisation iterations. Experimentation with the penalty function parameters within the LFOPC solver of Dynamic-Q did not have sufficiently noticeable effects on the inequality constraint's convergence history. This is because the LFOPC solver finds a feasible optimum of the optimisation approximate sub-problem for every iteration, regardless of the changes in the penalty function values. This can be attributed to the smooth nature of the spherically quadratic approximate objective and constraint functions. It is postulated that the complication arises from a poor approximation of the objective and constraint functions due to the high levels of numerical noise in the full simulation model, and or gradient information.

Closer inspection of the tyre deflection at very low damping values, indicated unrealistically high levels of tyre deflection. The high tyre deflection is

attributed to the linear vertical tyre stiffness used in the ADAMS model. The tyre damping was increased to help overcome this effect with no suitable improvement at double the measured tyre damping. The logical step is to implement a non-linear vertical tyre stiffness, to overcome the problem. This was not implemented due to time constraints as the current tyre model used in the MSC.ADAMS model cannot accommodate a non-linear vertical tyre stiffness. It is thus suggested that a non-linear vertical tyre stiffness should be implemented in the tyre model used, before a decrease in the high levels of noise associated with the tyre hop inequality constraints can be achieved.

Another suggested method of overcoming the noise levels present in the objective and constraint functions as a result of the simulation model, is by re-formulating the multi-body dynamics solver's convergence criteria. The implementation of the proposed method, however, requires access to the code, and for this reason was not implemented for this research.

7.4 Conclusions

This chapter investigated the optimisation of an off-road vehicle's suspension characteristics for ride comfort and handling, where the suspension characteristics are defined by numerous design variables.

The handling optimisation highlighted the design variables that have a predominant effect on the handling performance, but also that the other design variables do contribute to the improvement of the optimum objective function achievable. However, some variables showed poor sensitivity and needed to be rescaled to improve the sphericality of the optimisation problem. This highlighted, that ensuring that the design variables vary over the same range and have equal orders of magnitude, does not necessarily guarantee good convergence to the optimum.

The ride comfort optimisation, illustrated that erratic optimisation convergence histories can be a result of over sensitive design variables in

comparison to the rest of the design variables. These over sensitive design variables were identified and rescaled, resulting in greatly improved optimisation convergence history of the objective function. However, difficulty was encountered with satisfying the tyre hop inequality constraints. This difficulty could be as a result of the tyre model's use of a linear vertical tyre stiffness, leading to unrealistically high tyre deflections in the presence of low suspension damping. Increasing the vertical tyre damping did not result in a sufficient improvement. A non-linear vertical tyre stiffness should be implemented in the model in future.

This chapter provides the optimisation engineer with some valuable methods for identifying scaling problems in the definition of the optimisation problem. While all the complications associated with noise in the optimisation process have not been addressed, feasible suggestions for future work have been proposed.

Chapter 8

Automatic Scaling of Design Variables

With the difficulties encountered in Chapter 7, it is proposed that an automatic scaling methodology be implemented within Dynamic-Q. This should limit the number of investigations and time the optimisation engineer spends on the formulation of the optimisation problem, performing optimisation runs with poor convergence, and repeating the process. This automatic scaling methodology is proposed for unconstrained optimisation problems with only design variable upper and lower bounds. This scaling aims to improve the sphericality of the optimisation problem. Figure 8.1 illustrates the typical problem with an elliptic problem, whereby the optimisation is only very sensitive to one variable, and thus takes longer to reach the optimum point. The primary assumption with this methodology is that the design variables are uncoupled. While this is an oversimplification of the optimisation problem, from the surfaces generated in Chapter 5 this assumption is not far from the physical problem.

The basic proposed methodology can be summarised as follows:

1. Scale design variables using their upper and lower limits to between zero and one.

2. Perform one function evaluation at the middle of the design space

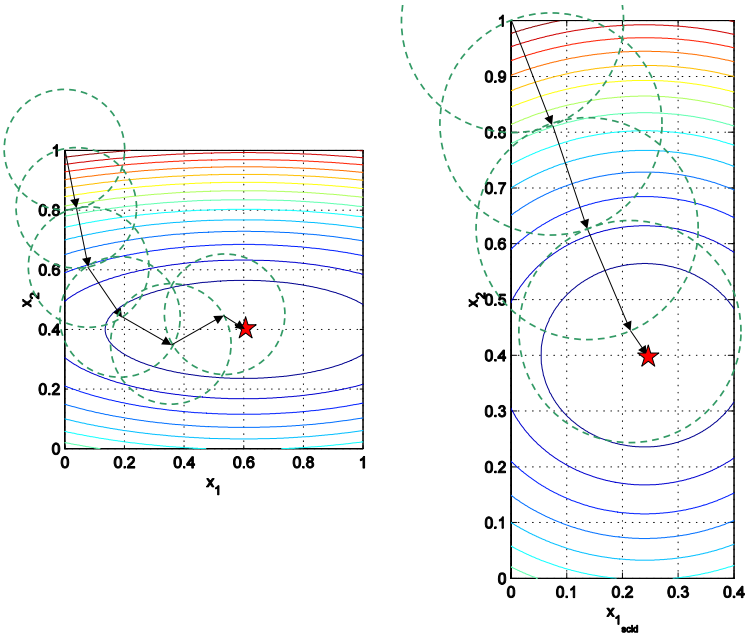


Figure 8.1: Illustration of the effect of ellipticity and sphericality on the convergence to the optimum

3. For each design variable perform 2 additional function evaluations with a perturbation of 50 % about the middle of the design space.

4. Construct a quadratic approximation of the objective function using the above simulation data, with respect to each design variable.

5. Rescale the scaled design variables so that the quadratic coefficient is equivalent to 1.

6. Perform the optimisation with the new rescaled design variables, but report the equivalent unscaled design variables at each iteration point.

This methodology should thus help to eliminate the ellipticity of the problem, if the design variables are approximately uncoupled. In the vehicle dynamics application the simplified models will be used for performing the scaling, however, under normal circumstances the cost of this scaling is equivalent to one iteration, i.e. $2n + 1$ function evaluations where n is the number of design variables, if central finite differencing is used for gradient information. The proposed automatic scaling will now be formally formulated.

8.1 Formulation of Unconstrained Automatic Scaling

Six steps need to be followed for the implementation of the automatic scaling methodology within Dynamic-Q.

Step 1: Scale the design variables x_i between their upper $\check{k}_i$ and lower $\hat{k}_i$ boundary values, so that the design variable ranges from zero to one, as follows:

$$z_i = \frac{x_i - \hat{k}_i}{\check{k}_i - \hat{k}_i} \tag{8.1}$$

where i ranges from 1 to the number of design variables n.

Step 2: Perform $2n + 1$ function evaluations to obtain function values for the construction of the quadratic approximations. This is known as central composite design (CCD). The function evaluations are performed with a 50% perturbation from the middle of the design space. The initial function evaluation $f(\mathbf{x}_{mid})$ is at the middle of the design space (mid-space). The next function evaluations are performed with the design variables at the mid-space value $\mathbf{x}_{mid}$, but a perturbation in only the i^{th} design variable. Thus the function evaluations can be defined as:

$$\hat{f}_i = f(\mathbf{x}_{mid}, \hat{x}_i) \check{f}_i = f(\mathbf{x}_{mid}, \check{x}_i) \tag{8.2}$$

where $\hat{x}_i$ corresponds to the equivalent $\hat{z}_i$ which is defined as:

$$\hat{z}_i = z_{0_i} - 0.5 \tag{8.3}$$

and $\check{x}_i$ corresponds to the equivalent $\check{z}_i$ which is defined as:

$$\check{z}_i = z_{0_i} + 0.5 \tag{8.4}$$

Thus the objective function is evaluated with a 50% perturbation on either side of the mid-space point $\mathbf{x}_{mid}$, resulting in the whole design space being approximated. All the optimisation up until now has been done with a 3% perturbation on either side of the current iteration point for the evaluation

of the central finite difference gradient. It is decided that by considering the whole design range a suitable evaluation of the effective curvature of the optimisation problem is obtained, without being dramatically affected by numerical noise. This then ensures that the sensitivity of the design variables over the whole design space is taken into account.

Step 3: Use the above determined objective function values to construct approximate quadratic approximations of the objective function with respect to the design variable as follows:

$$\tilde{f}(z_i) = a_i {z_i}^2 + b_i z_i + c_i \tag{8.5}$$

Step 4: Rescale the design variables $\mathbf{z}$ so that the corresponding a_i term of the approximated objective function will be 1. The rescaled design variable will thus be defined as:

$$X_i = z_i t_i \tag{8.6}$$

where the design variable scale factor t_i is defined as:

$$t_i = \sqrt{a_i} \tag{8.7}$$

This scaling, however, has the problem that it tends to zero when a_i tends to zero, and X_i tends to infinity when a_i tends to infinity. It is thus proposed that should a_i tend to zero, then the quadratic approximation tends to a straight line, and that this straight line should have a gradient of + or - 1, this means that the b_i term should be used for the scaling. If, on the other hand, a_i becomes very large the scaling will result in a very large design space with respect to that design variable, thus it is proposed that the upper limit of the design space/variable range should be 20, corresponding to an a_i value of 400. The lower limit to the design space is chosen as 0.2, corresponding

to an a_i value of 0.04. The following if loop then applies:

$$\begin{aligned} &if \quad a_i \leq 0.04 \\ &t_i = |b_i| \\ &elseif \quad a_i \geq 400 \\ &t_i = \sqrt{400} = 20 \\ &end \end{aligned} \tag{8.8}$$

Again the problem of tending to infinity when $|b_i|$ tends to infinity, exists. Should $|b_i|$ tend to zero, then t_i also tends to zero. Thus the additional if statement must be inserted:

$$\begin{aligned} &if \quad |b_i| \leq 0.2 \\ &t_i = 0.2 \\ &elseif \quad |b_i| \geq 20 \\ &t_i = 20 \\ &end \end{aligned} \tag{8.9}$$

The rescaled design variables X_i will thus be limited to the following:

$$0.2z_i \leq X_i \leq 20z_i \tag{8.10}$$

Step 5: Change the move limit so that it is still representative for the rescaled problem. The new move limit DM_n is a function of the number of design variables n, the original move limit dml, and the scale factors t_i, and is defined as follows:

$$DM_n = dml\sqrt{\frac{\sum_{i=1}^{n} {t_i}^2}{n}} \tag{8.11}$$

Step 6: Perform the optimisation with the rescaled design variables $\mathbf{X}$, and new move limit DM_n, but report the actual design variable values to the user. The Dynamic-Q design variables $\mathbf{X}$ will thus be converted for printout to the users design variables $\mathbf{x}$ as follows:

$$x_i = \frac{X_i}{t_i}(\check{k}_i - \hat{k}_i) + \hat{k}_i \tag{8.12}$$

8.2 Concept Test

This automatic scaling methodology is first tested on simple severely elliptic, analytical design problems. The first test is using a two design variable analytic problem described by the objective function:

$$f(\mathbf{x}) = {x_1}^2 + 10{x_2}^2 \tag{8.13}$$

Were the objective function $f(\mathbf{x})$ is uncoupled with respect to x_1 and x_2. Upper and lower bounds on the design variables are defined as:

$$-1 \leq x_{1,2} \leq 1 \tag{8.14}$$

and an initial starting point of [1 1]. The gradient information was determined analytically, and the performance of the standard form of Dynamic-Q (Snyman and Hay 2002) was compared to the automatic scaling version of Dynamic-Q to be known as Ascl-Dyn-Q. Figure 8.2 illustrates the comparison of the convergence histories in the design space for the standard form of Dynamic-Q and Ascl-Dyn-Q. It is observed that Ascl-Dyn-Q moves much faster towards the optimum. The function error is determined as defined by Snyman and Hay (2002) as:

$$f(err) = \frac{\|f_{act} - f^*\|}{1 + \|f_{act}\|} \tag{8.15}$$

For the above optimisation problem the results are tabulated in Table 8.1, line *cp* 1.

The second problem *cp* 2 is a skew problem, where the dependence of $f(\mathbf{x})$ with respect to x_1 is coupled to x_2, described by the objective function:

$$f(\mathbf{x}) = {x_1}^2 + 10{x_2}^2 + 3x_1x_2 \tag{8.16}$$

With upper and lower bounds on the design variables defined as:

$$-1 \leq x_{1,2} \leq 1 \tag{8.17}$$

and an initial starting point of [1 1]. It can be seen from Figure 8.3 that the Ascl-Dyn-Q algorithm moves to the optimum in the same manner as the

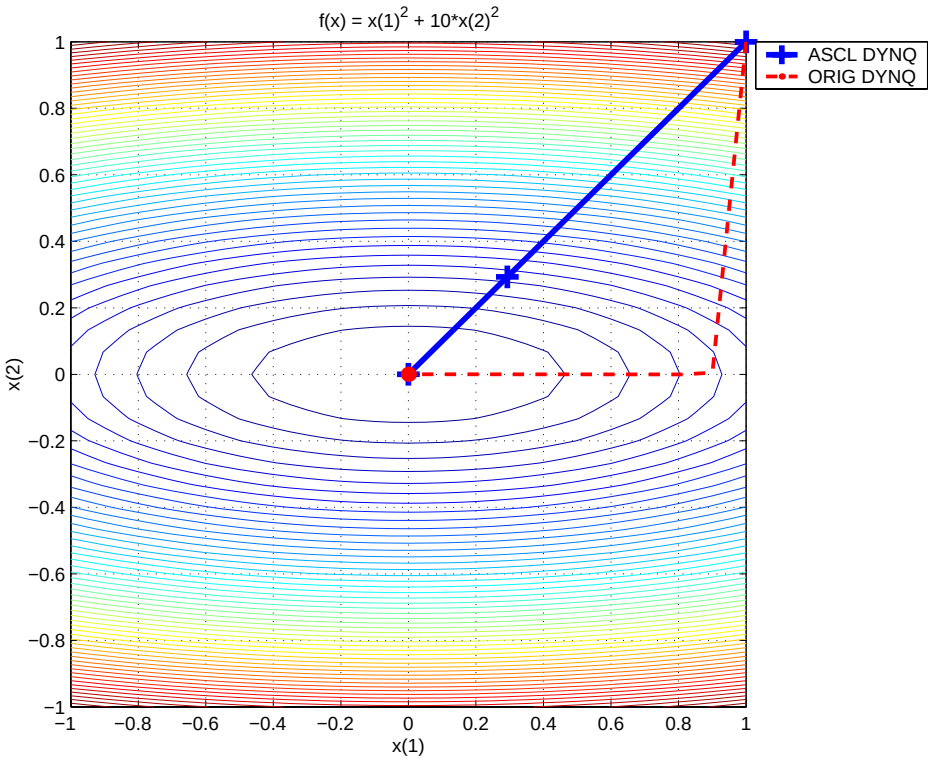


Figure 8.2: Comparison of standard Dynamic-Q convergence to optimum and Dynamic-Q with automatic scaling, for test 1

standard algorithm, but approaches the optimum from the other side. This is the limit of the permissible cross-coupling of design variables, where the algorithms exhibit almost equal performance, *cp* 2 in Table 8.1.

8.3 Modification for Constrained Problems

With the success achieved with the proposed automatic scaling procedure, the methodology was expanded to include constrained optimisation problems. It is proposed that, because Dynamic-Q makes use of LFOPC for the optimisation of the approximate sub-problem at each iteration, the constraints could be included by using the penalty function approach, as in LFOPC. The resulting penalty function should thus be made spherical, as opposed to just the underlying objective function. LFOPC solves the penalty function in a three part approach (Snyman 2000), thus the question must be asked as to what penalty parameter multiplication factor should be used. LFOPC first solves the approximate sub-problem using a low penalty function multiplication

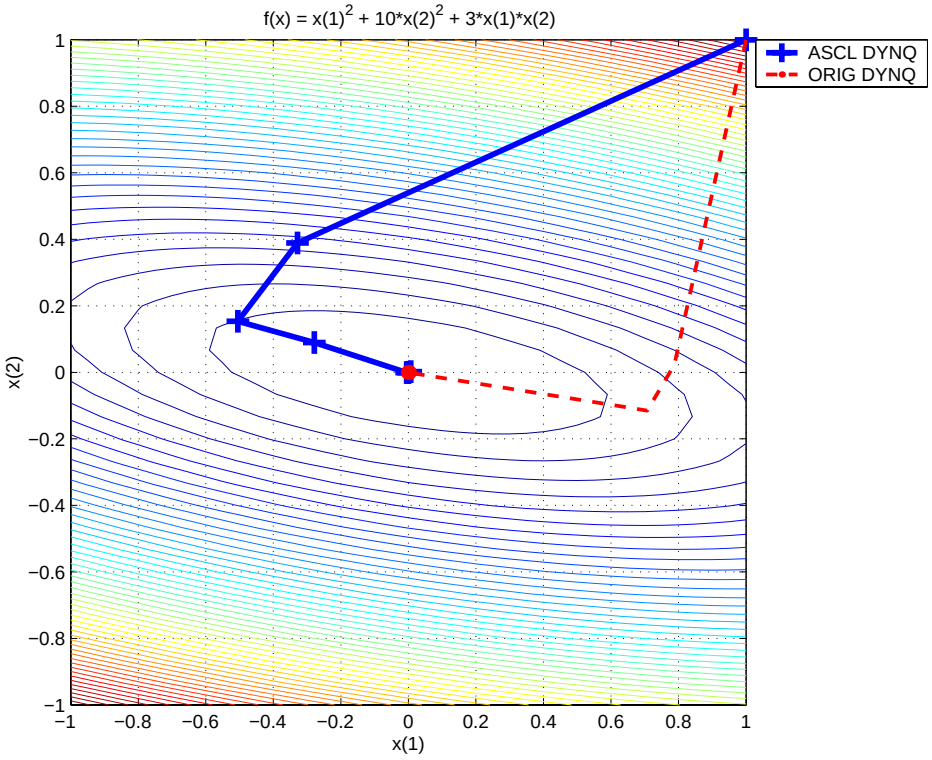


Figure 8.3: Comparison of standard Dynamic-Q convergence to optimum with Dynamic-Q with automatic scaling, for test problem 2

factor, and then increases the penalty function multiplication factor in the next phases. It is thus proposed that the violated constraint should be added to the objective function value using the lowest multiplication factor. The quadratic approximations with respect to each design variable are then fitted to the resulting penalty function. Where the penalty function is defined as follows:

$$\underset{w.r.t.x}{minimize} \quad P(\mathbf{x},\mu) \tag{8.18}$$

where

$$P(\mathbf{x},\mu) = f(\mathbf{x}) + \sum_{j=1}^{m} \mu_j g_j(\mathbf{x}) + \sum_{j=1}^{r} \mu_j h_j(\mathbf{x}) \tag{8.19}$$

where the penalty multiplier μ_j is defined by the if statement for $g_j(\mathbf{x})$ as:

$$\begin{gathered} if \quad g_j(\mathbf{x}) \leq 0 \\ \mu_j = 0 \\ else \quad \mu_j >> 0 \\ end \end{gathered} \tag{8.20}$$

Table 8.1: Results for the standard Dynamic-Q and Auto-Scaling Dynamic-Q methods

Problem #	n	$f(act)$	Dynamic-Q			Ascl Dyn-Q		
			# iter	f^*	$f(err)$	# iter	f^*	$f(err)$
cp 1	2	0.00e+00	7	1.53e-12	1.53e-12	3	8.99e-11	8.99e-11
cp 2	2	0.00e+00	7	1.70e-12	1.70e-12	8	4.24e-11	4.24e-11
Hock 2	2	5.04e-02	7*	4.94e-00	error	11*	4.94e-00	error
Hock 13	2	1.00e+00	6	9.99e-01	1.00e-08	6 nc	1.00e+00	3.00e-07
Hock 15	2	3.07e+02	17	2.13e+02	4.35e-01	14	error	error
Hock 17	2	1	16	1	< 1.00e-08	10 nc	1	< 1.00e-08

* - converged to local minimum

nc - no constraints considered with scaling

and the if statement for $h_j(\mathbf{x})$ as:

$$\begin{aligned} &if \quad h_j(\mathbf{x}) = 0 \\ &\mu_j = 0 \\ &else \quad \mu_j >> 0 \\ &end \end{aligned} \tag{8.21}$$

Some random test problems, of the ones on which Dynamic-Q was tested and presented in Snyman and Hay (2002) are resolved using the auto-scaling methodology. The test problems are from the book of Hock and Schittkowski (1981), and given in Appendix A for the readers convenience.

From the results it can be seen that the auto-scaling improves the optimisation convergence to the optimum, for most test cases, except badly skew elliptic problems like the Rosenbrock problem. From the results the feasibility of using the penalty function for constrained optimisation problems has not been shown, as there are difficulties as to what value to use for the penalty function multiplication factor.

With the success of the automatic scaling a search was done for similar

novel approaches to scaling of the design variables. Most researches suggest the normalisation of the design variables to range between zero and one, as in Snyman (2005b) and Lasdon (2001). Willcox (2006) suggests inspecting the Hessian matrix at the converged optimum design point. The condition number of the Hessian matrix is evaluated. If the condition number is greatly larger than one, the matrix is ill-conditioned, and the design variables are transformed linearly to minimise the condition number of the solution. This is, however, only performed after the optimisation algorithm has converged to a solution, while the scaling proposed in this thesis is done over the whole design space at the begining of the optimisation process. This ensures that the global problem is scaled to be more spherical. The other advantage of the automatic scaling suggested here, is that the Hessian matrix does not need to be constructed, greatly reducing gradient evaluations, that are normally very costly in typical engineering problems.

If the Hessian needs to be calculated in order to better scale the design variables, Danchick (2006) suggests an efficient and accurate method for computation of the Hessian matrix. Danchick makes use of central finite difference quotients and extrapolation-to-the-limit to achieve a h^4 level of accuracy, where h is the finite difference step size. The computational cost is $2n(n+1)+1$ function evaluations per Hessian matrix evaluation, where n is the number of design variables. This is then implemented in an optimisation algorithm that makes use of Hessian decomposition and eigenvalue shifting to follow a ridge in a difficult skew optimisation problem like Rosenbrock's parabolic valley. The determination of this Hessian matrix should be considered over the whole design space before the optimisation process should be considered in the future, but will probably not be necessary for most well defined engineering problems.

8.4 Implementation in the Vehicle Suspension Problem

The automatic scaling method proposed above was implemented for the 14 design variable optimisation, and the results were compared to the those obtained with Dynamic-Q without automatic scaling. Because of the apparent hopping about an optimum point, the convergence histories presented are in the form of the best feasible point at the current iteration point. This approach is borrowed from the genetic algorithm and particle swarm community. If an improved solution is not achieved after a certain number of iterations, the optimisation is terminated. The number of iterations before termination is, however, difficult to select, as this may result in premature termination. It was also determined that the optimisation should not be permitted to terminate within the first 10 iterations.

Presented in Figure 8.4, is the comparison of the optimisation convergence histories for optimisation with automatic scaling (ascl) and without (std). It is observed that the automatic scaling terminates at a better optimum than without automatic scaling, and the design variables do not get stuck in local minima as for the standard optimisation. This local minimum is used to start the optimisation using automatic scaling in order to achieve a better optimum, the results are presented in Figure 8.5. It can thus be concluded that the automatic scaling was successful for the optimisation of 14 design variables for handling.

For ride comfort, the decision of when to terminate the optimisation will impact on the performance of the optimisation methods. Presented in Figure 8.6 is the optimisation convergence histories for the standard form of Dynamic-Q (std) and with automatic scaling implemented (ascl). The use of the penalty function for the automatic scaling as discussed in section 8.3, was not used here, due to difficulties associated with the correct magnitude of the penalty multiplier. It can be seen that the automatic scaling reaches a better

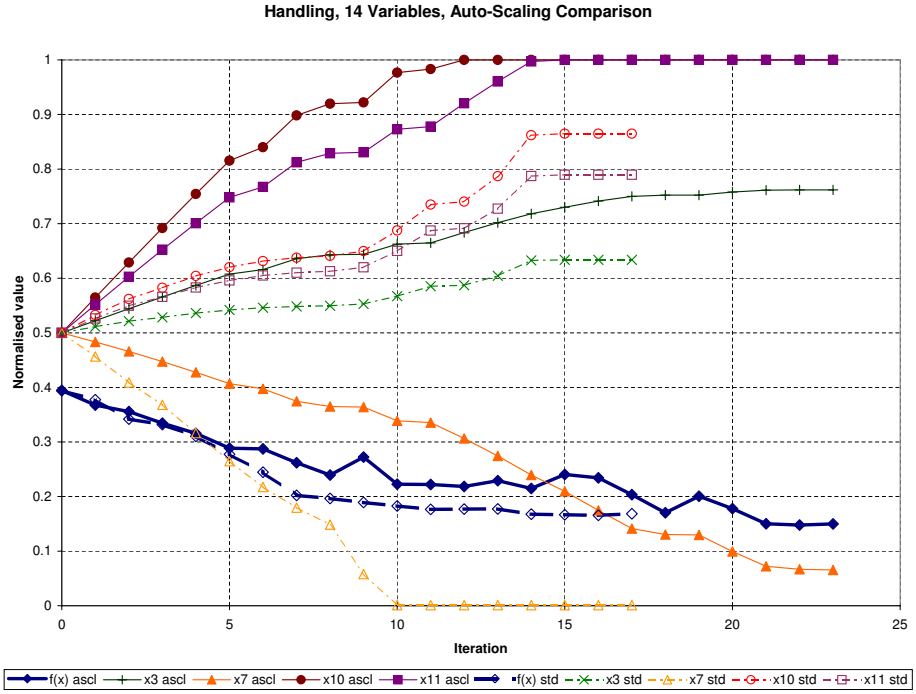


Figure 8.4: Comparison of convergence histories for standard Dynamic-Q and for the implementation of the automatic scaling, for 14 design variable handling

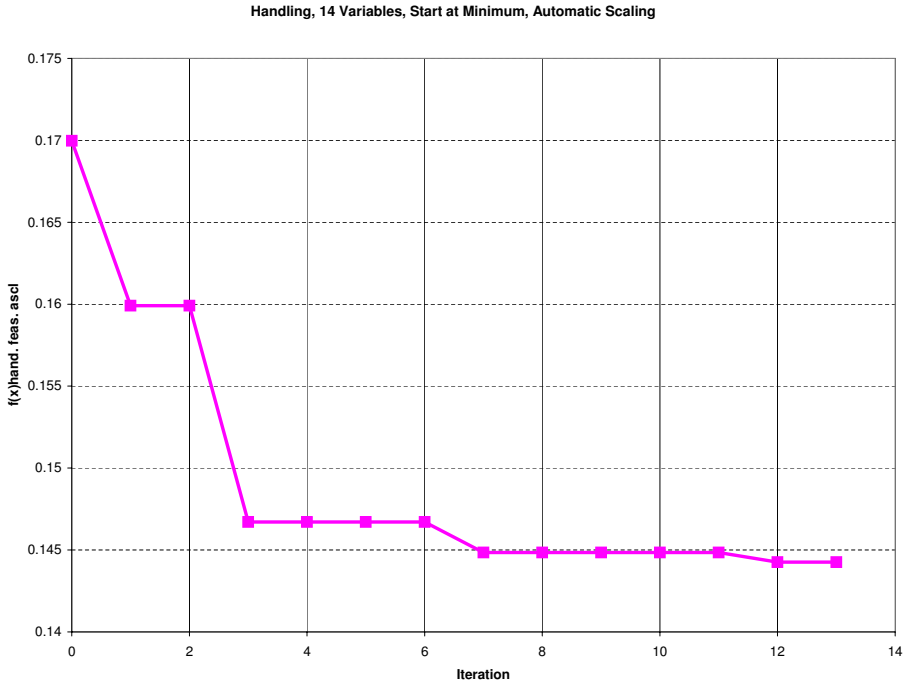


Figure 8.5: Optimisation convergence history for handling, starting at the found minimum

minimum than the standard form, however, this is not the optimum point, as better optima were reached using only 4 design variables (see Chapter 6, Figure 6.9). The minimum point achieved in Figure 8.6, for the automatic scaling (f(x) feas. ascl, iteration 18) was then used as a starting point for the optimisation of the ride comfort, with the design variables subjected to a 30% range about the minimum found in Figure 8.6. The optimisation convergence history is presented in Figure 8.7. It is observed that an equal minimum is reached as for the 4 design variables.

From the optimisation results the optimum damper characteristics for handling (Figure 8.8) and ride comfort (Figure 8.9), are presented. It is also found that for optimum handling the static gas volume must be 0.1 liter, while for optimal ride comfort the static gas volume in front should be 0.39 liter, and at the rear 0.46 liter. The handling gas volume thus ran to the lower boundary, but the ride comfort static gas volume did not. This can be attributed to the tyre hop inequality constraints. The optimal driver vertical RMS acceleration is 1.1 m/s^2 and the rear passenger vertical RMS acceleration is 1.1 m/s^2. The optimal body roll velocity RMS value is 0.52 $^o/s$ and the maximum roll angle is 3.1 ^{o}.

8.5 Conclusions

The automatic scaling methodology was proposed, and implemented on several analytic problems with success. Automatic scaling was then applied to the vehicle dynamics problem of numerous design variables.

The automatic scaling methodology was implemented with success. The optimal handling and ride comfort were determined, where it was found that the handling setting would require a small gas volume, and stiff dampers front and rear, while the ride comfort required soft front and even softer rear damping, and a static gas volume of 0.39 liter in front and 0.46 liter at the rear.

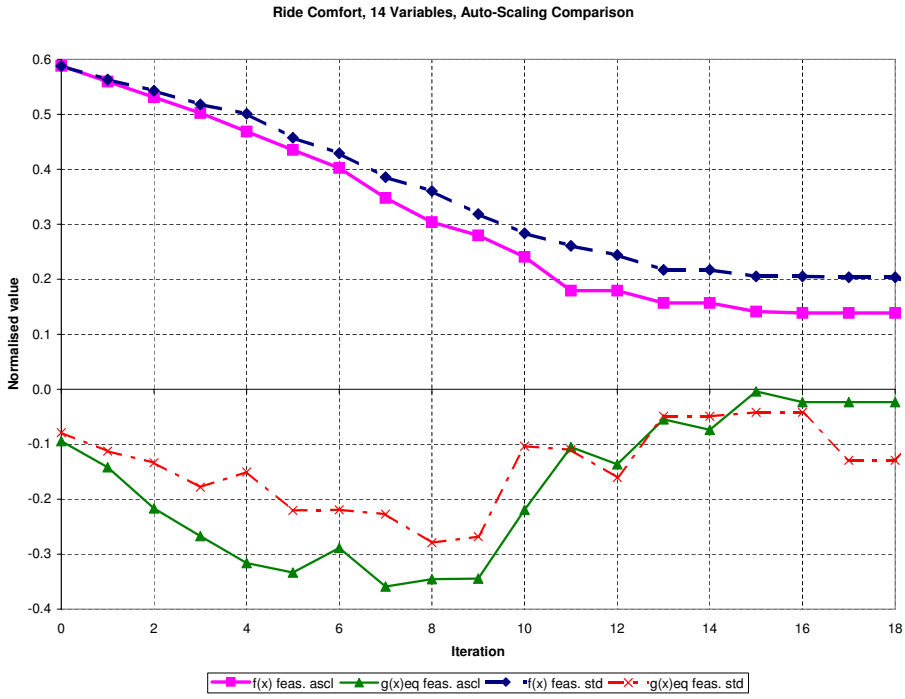


Figure 8.6: Comparison of optimisation convergence histories for standard Dynamic-Q and for the implementation of the automatic scaling (14 design variable ride comfort)

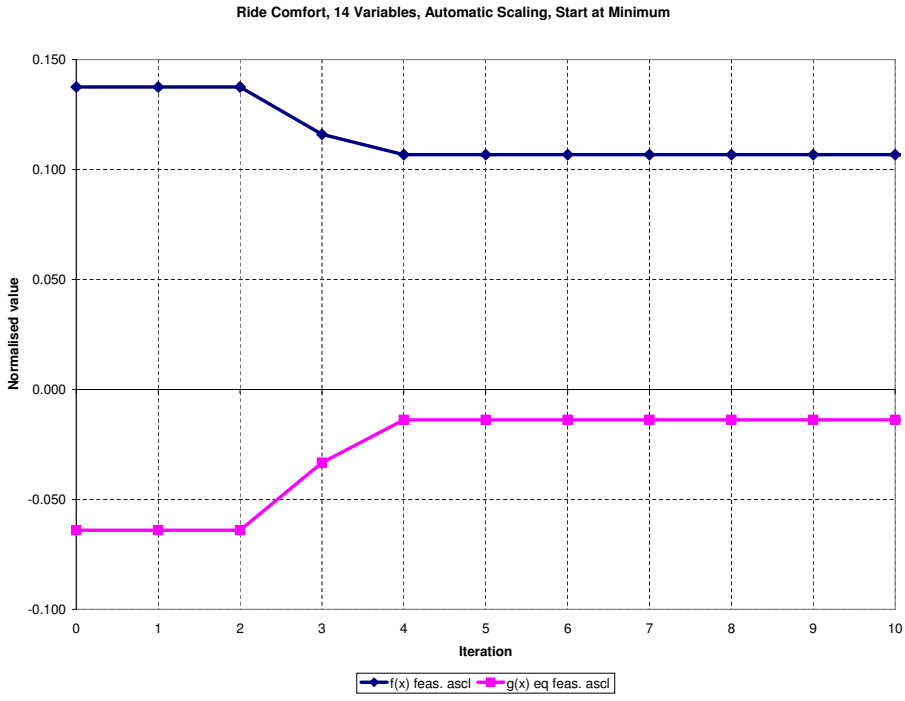


Figure 8.7: Optimisation convergence history for starting at minimum, using automatic scaling

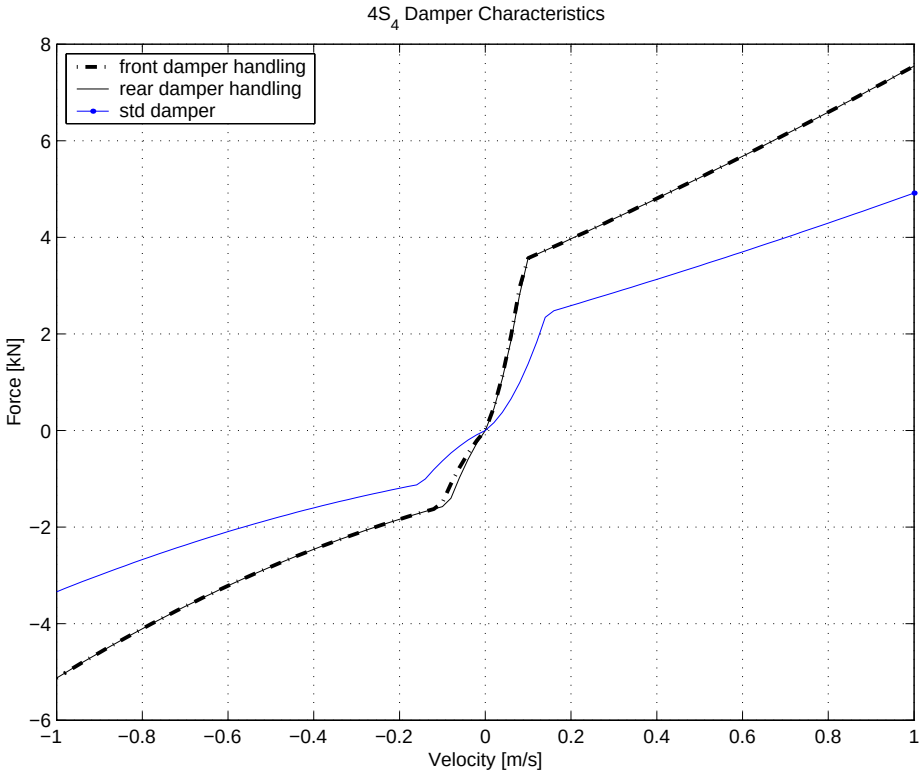


Figure 8.8: Optimum damper characteristics for handling compared to the baseline rear damper

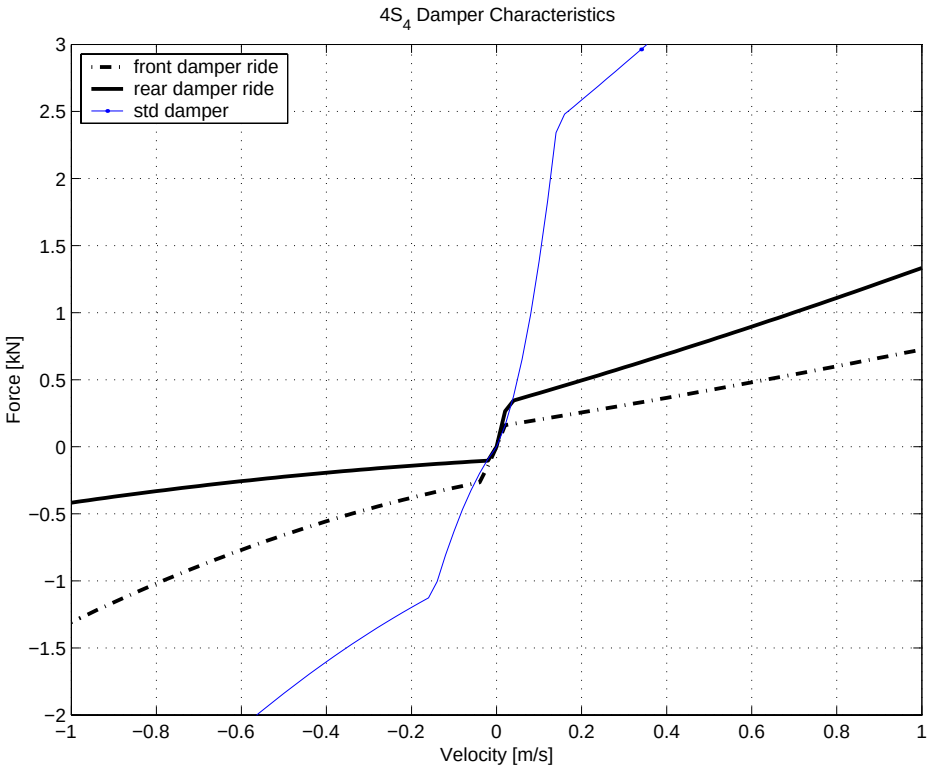


Figure 8.9: Optimum damper characteristics for ride comfort

The automatic scaling methodology can be further improved with the investigation of the Hessian matrix. This Hessian matrix can then be transformed so as to better scale skew problems.

Further investigation into the weight of the penalty function multiplication parameters is needed, so that scaling of constrained problems can be improved.

Chapter 9

Combined Optimisation

With the use of simplified models for gradient information validated, the models are combined to represent the vehicle performing a handling manœuvre on a rough terrain. For the combined ride comfort and handling optimisation, the vehicle performs the double lane change over the Belgian paving. The full simulation model is used, as before, once per iteration for the exact objective function values and constraint values. The Matlab models remain the same. However, the ride model will be used to observe the ride dynamics gradient tendencies, and the handling model for the handling dynamics gradient tendencies. This work was performed before the proposal of the automatic scaling methodology. A study was conducted as to how best to consider the optimisation of the compromise passive suspension. This is done to determine the methodology needed when including the control strategy of the $4S_4$ system for optimisation.

9.1 Handling Followed by Ride Comfort

First the vehicle will be optimised for handling, subject to the tyre hop inequality constraints, and then optimised for ride comfort starting from the point where the handling optimisation converged, for two design variables. The ride comfort is optimised subject to the tyre hop inequality constraints, and an additional inequality constraint that the optimised handling $f^*(\mathbf{x})_{hand}$ may not decrease by more than 20% (compared to the optimised handling

result) as stated below:

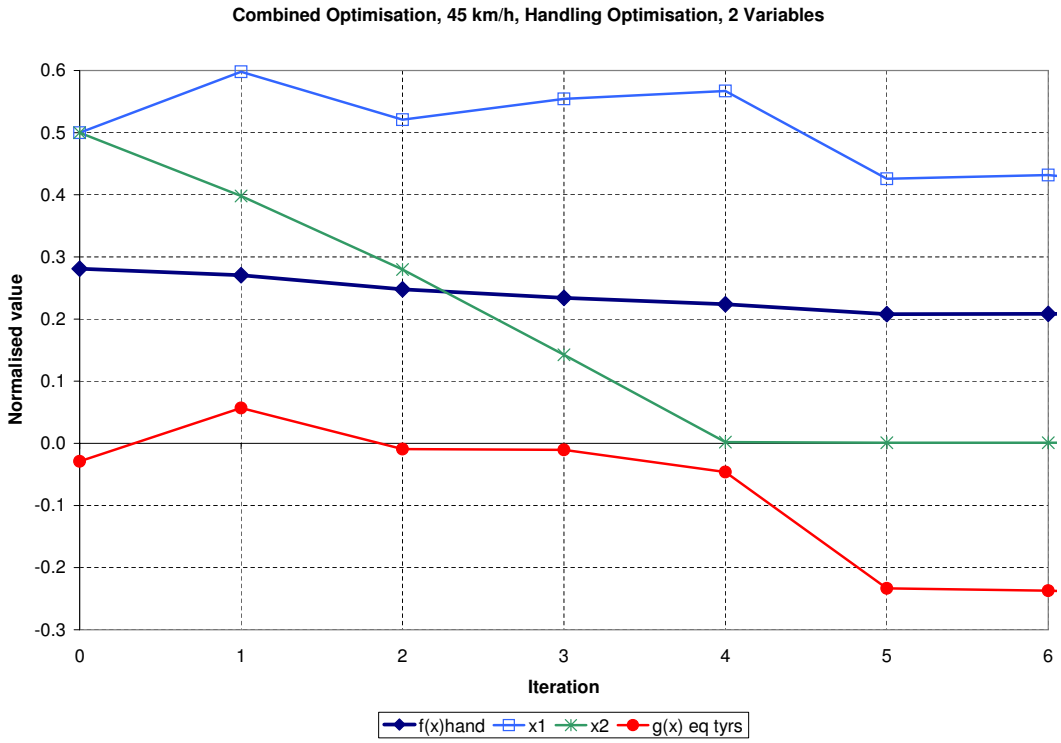
$$g(\mathbf{x})_{hand} = 10(f(\mathbf{x})_{hand} - 1.2f^*(\mathbf{x})_{hand}) \leq 0 \tag{9.1}$$

The 20% parameter was selected as it was found that for optimisation runs where the handling constraint was 5 or 10 %, the handling constraint could not be satisfied, if improvements in ride comfort were achieved. The value of 20% was thus found to be a reasonable constraint value. This value would, however, typically depend on the design requirements for the specific vehicle being optimised. The multiplication by 10 was used to better normalise the constraint values between -1 and 1.

The optimisation convergence history for two design variables is presented in Figure 9.1. The equivalent tyre hop constraint is plotted as defined in equation (6.2). The top graph refers to the handling optimisation where the objective function is defined as in equation (5.7), and the bottom graph is for the ride comfort optimisation, where the objective function is defined as in equation (5.6). It can be seen that the optimisation convergence history is well behaved for the handling optimisation, and results in an objective function value of approximately 0.21, which is equivalent to a body roll angle of 3 ^{o}, and a RMS body roll velocity of 1.3 $^o/s$. The ride comfort optimisation, subjected to the handling constraint, has a poorly behaved convergence history, and does not converge to a clear optimum. If iteration 18 is considered as the best minimum, the driver RMS vertical acceleration is approximately 2.2 m/s^2, which is considered as extremely uncomfortable (Els 2005), and needs to be improved. The ride comfort can be greatly improved but at the expense of handling.

9.2 Maximum of Ride Comfort and Handling

The results for handling followed by ride comfort optimisation, prompted the investigation into using the maximum value of the four normalised objective function parameters (roll angle, RMS roll velocity, driver comfort, passenger comfort) as the objective function value. The objective function is thus



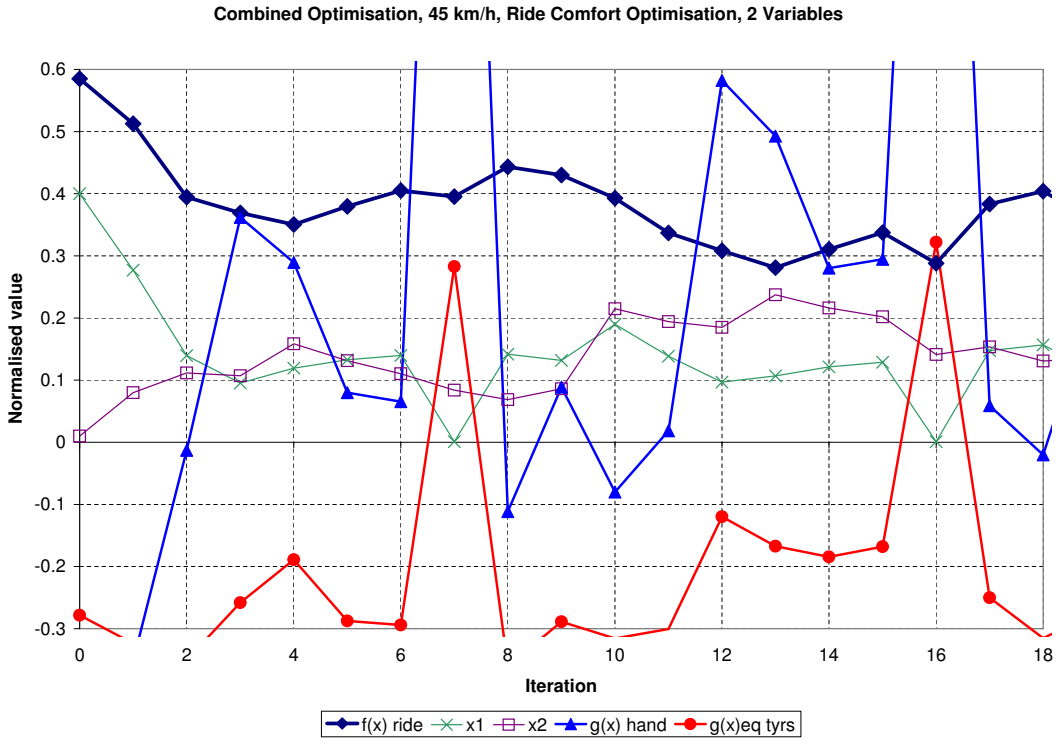


Figure 9.1: Combined convergence history, first handling optimisation, then ride comfort.

defined as follows:

$$f(x) = max(f(x)_{hand}, f(x)_{ride}) \tag{9.2}$$

The disadvantage of the nature of this objective function is the now inherent discontinuities due to the maximum function. However, very reasonable results were achieved as illustrated by Figure 9.2. In the figure, $f(x)hand$ is the handling objective function value as defined by equation (5.7), $f(x)ride$ is the ride comfort objective function as defined by equation (5.6), and the equivalent tyre hop constraint $g(x)eq$ defined by equation (6.2). Additionally it is observed that the overall optimum is the equalization of the two objectives. When considering the final design configuration, iterations 3, 6 and 9, are repeated identical minima, and should be considered for the acceptable band of the design variables, to return objective function values of approximately 0.32. This results in vertical RMS accelerations of approximately 1.8 m/s^2, body roll angle of 4^o, and a RMS roll velocity of 1.9 $^o/s$. The optimisation convergence took fewer iterations than the optimisation of handling followed by ride comfort, even though the objective function is of a discontinuous nature, due to the maximum function.

The use of the maximum function for the objective function was expanded to four design variables, and started in the same place as for two design variables, namely the middle of the design space. The results, presented in Figure 9.3, illustrate the excellent convergence to the optimum, of identical magnitude as for two design variables, but the design variable values differ. Although it is evident that multiple local minima exist, the optimisation converges to identical objective function value minima.

With the difficulty encountered with the definition of ride comfort as a constraint and optimising handling, yet excellent convergence history when using an equal weight of the two objectives, in the form of the maximum function, a pareto front will now be constructed, between the handling and ride comfort objective functions.

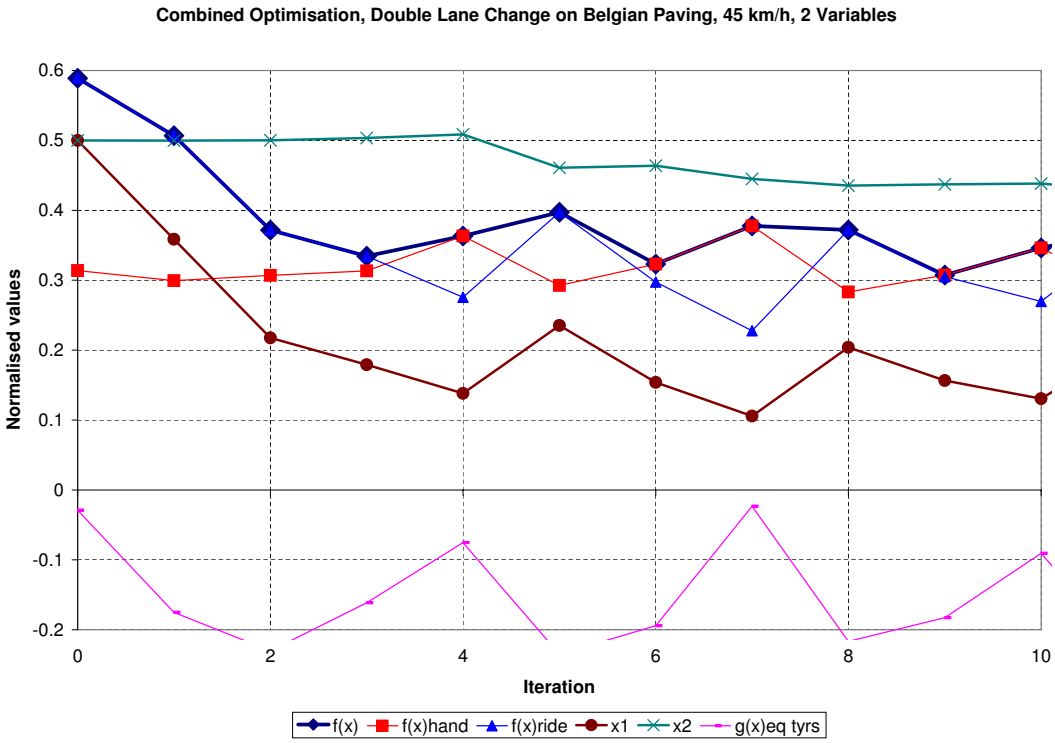


Figure 9.2: Combined optimisation convergence history, maximum of handling and ride comfort objectives, 2 design variables.

9.3 Pareto Optimal Front

With the success of the optimisation results, but the vastly varying optimal design points in the design space, the simplified model was used to investigate the trends in terms of the pareto optimal front of feasible points for ride comfort and handling, subject to the tyre hop constraints. Random points in the four design variable space were generated and their objective and constraint function values evaluated. This would traditionally give the design engineer the necessary insight into which optimal suspension settings to select for a desired combination of ride comfort and handling. However, as shown in Figure 9.4, the random feasible points lie greatly inward of the pareto optimal front. Optimisation runs were performed where the objective function was defined as a weighted sum of the handling and ride comfort objective functions defined in Chapter 5. Figure 9.4, illustrates the optimisation convergence histories of the differing weighted objective

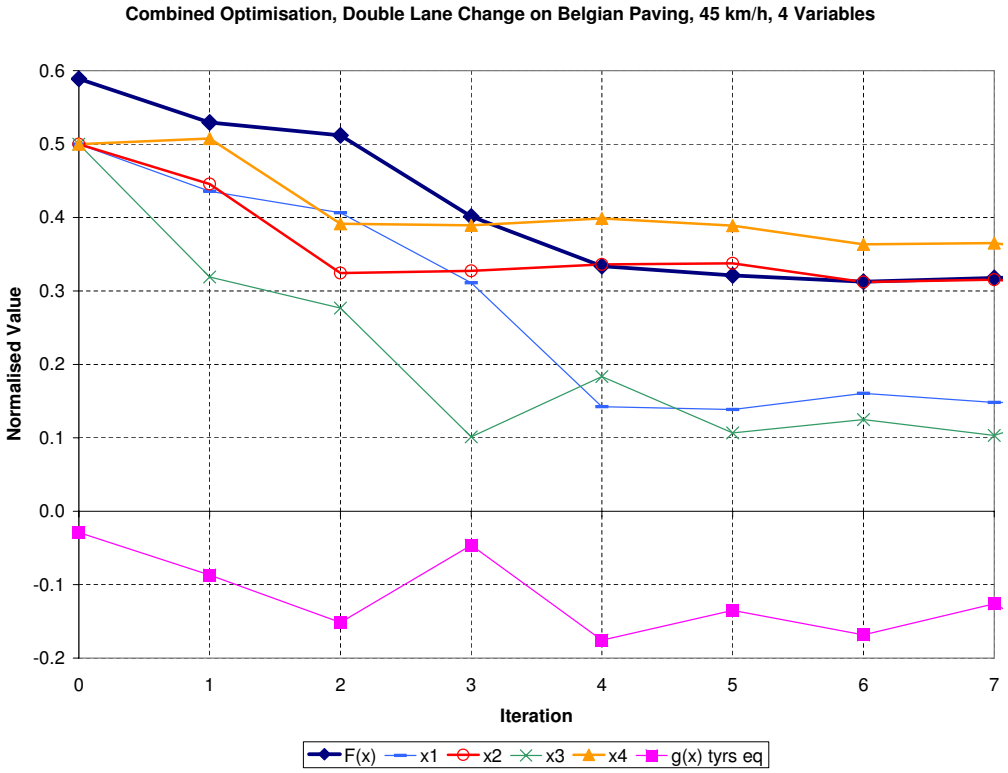


Figure 9.3: Combined optimisation convergence history, maximum of handling and ride comfort objectives, 4 design variables.

functions, to the pareto optimal front. The shortest distance from the pareto front to the zero point is generally accepted as the best compromise, however, this depends on which objective is most important to the vehicle being designed. In the case of the SUV, handling is a safety critical component, as these vehicles at their handling limit roll over before they slide out.

With the converged optimal points of Figure 9.4, the pareto optimal front and design variable values were plotted in Figure 9.5. From Figure 9.5, the change in the design variable with respect to a change in the ride comfort and handling objective function values can be quantified. The design engineer can now use this information to obtain a first order estimate as to the optimal design variable combination in order to achieve a desired point on the optimal pareto front. From Figure 9.5, it is observed that the rear suspension (design variables x_3 and x_4) have the greatest sensitivity on the ride comfort objective function value, when close to the handling optimum (i.e. f($\mathbf{x}$) handling $<$

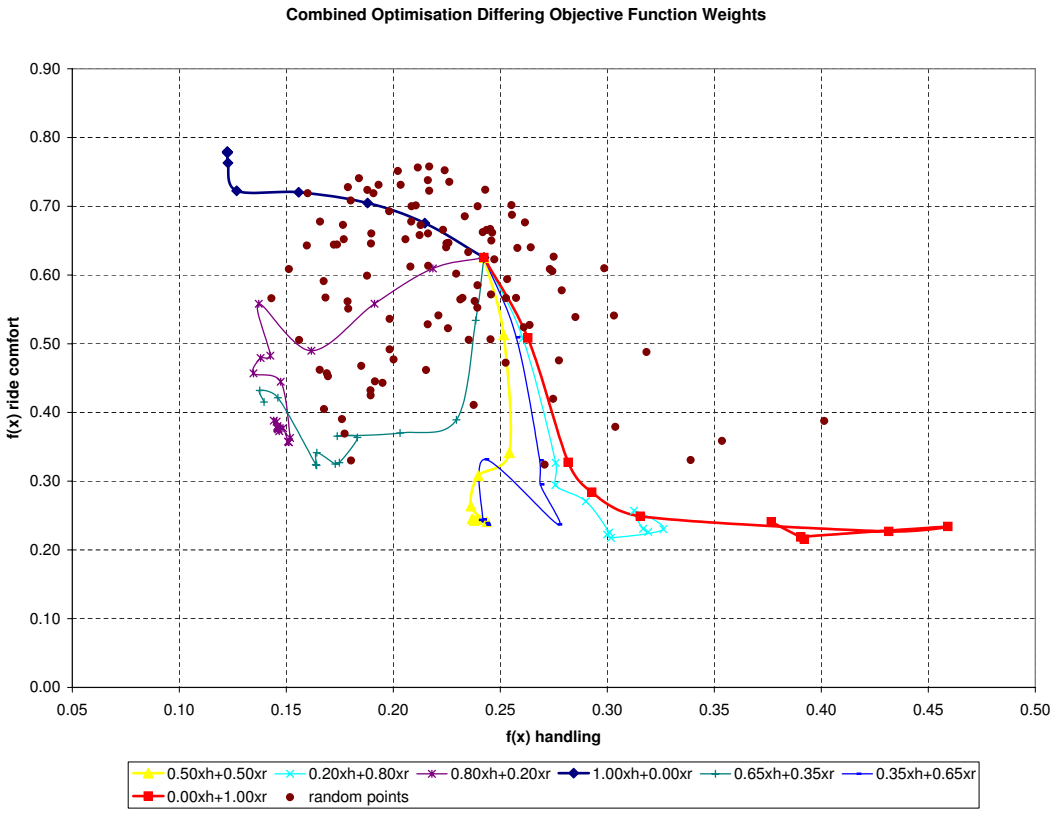


Figure 9.4: Investigation of convergence history to pareto front for different weights of the objective function for handling (h) and ride comfort (r), compared to random feasible points of design space

0.16). A large change in the value of design variables x_1, x_3 and x_4 will result in a dramatic improvement of the ride comfort, but a much smaller decrease in the handling objective function value, from the optimal to worst handling configuration. Also noticeable is the fact that the front gas volume design variable x_2, must be at it's stiffest setting (value of 0) for good handling, yet the rear gas volume design variable x_4 can be as large as 50% (value of 0.3) of the optimal handling gas volume.

From the pareto optimal front results presented in Figure 9.5, it can be concluded that the most feasible compromise point in the design space is for a handling objective function value of 0.16, and a ride comfort objective function value of 0.32. The use of the weighted objective functions to obtain the pareto optimal front is of importance, when designing the vehicle's

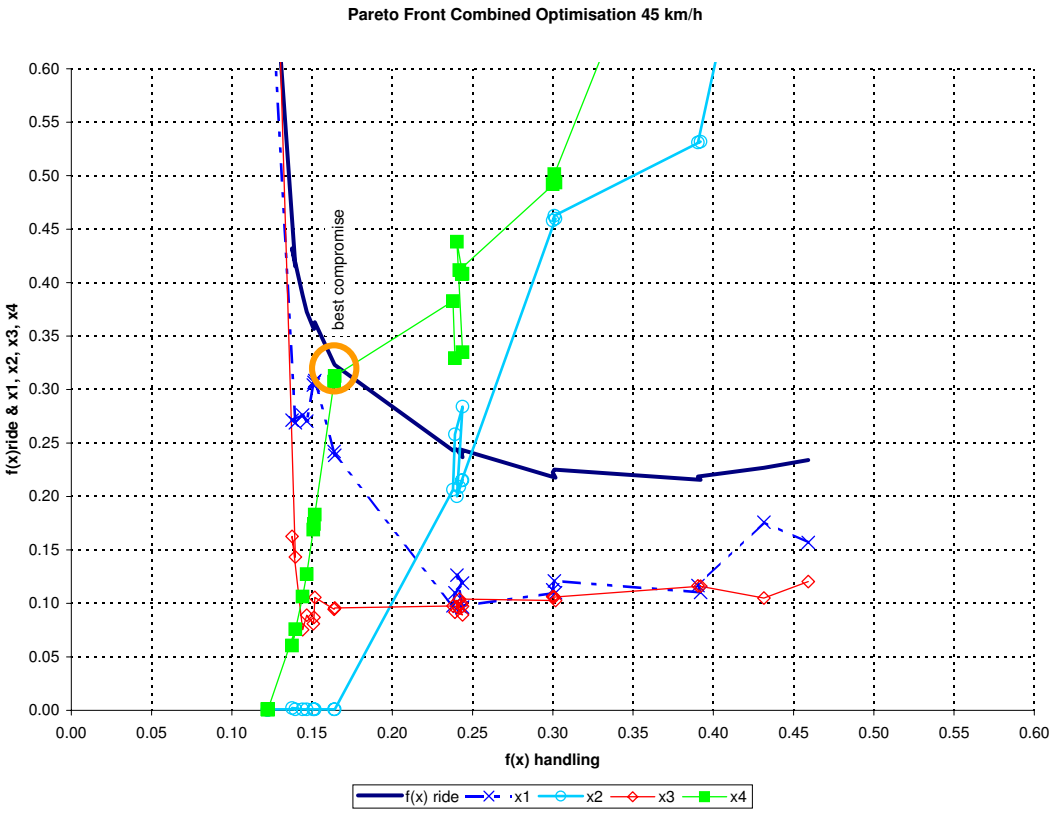


Figure 9.5: Pareto front plot including the change in the design variables along the pareto front

suspension system for the compromise ride comfort vs. handling setup. The necessary insight into the design variables most sensitive to improving the ride comfort with minimal loss of handling ability was obtained.

9.4 Summary of Results

Presented in Table 9.1 are the results for the optimisation runs. From the results it can be seen that the combined optimisation is a compromise between handling and ride comfort, especially when considering the use of the maximum function for the objective function. If reasonable handling is to be achieved, then the ride comfort suffers, while if good ride comfort is to be achieved then the handling suffers. This is the traditional compromise, that the $4S_4$ suspension avoids due to the ability to switch between the optimum handling and ride comfort settings. The resulting optimal damping

multiplication factors and spring gas volumes are presented in Table 9.2. Also noticeable when observing the parameters of the combined optimisation, is that the gas volume lies in the middle of the design space at 0.3 l, but that the damping should be 50% of the current baseline characteristic. This however, severely affects the handling stability of the vehicle as can be observed by the higher RMS roll velocity value. The most feasible compromise suspension setup was, however, achieved when considering the pareto optimal front.

The pareto optimal front provided the necessary insight into the problem in order to select the most feasible compromise. The resulting optimal damping multiplication factors and spring gas volumes were evaluated using the full MSC.ADAMS simulation model. It was found that the pareto optimal front objective function values were optimistic when compared with the actual full simulation model's objective function values. The full simulation pareto front values for four design configurations along the pareto front were evaluated and presented in Table 9.3. From the results it is observed that the pareto front displayed accurate trends, but with optimistic objective function values. From the pareto test points it can be seen that design configuration 3 returns acceptable ride comfort with a average decrease in handling of 31% over test point 1. This pareto methodology can thus in future be used to optimise a controllable suspension with included control system. Only the simplified models are optimised with objective functions being defined by differing weights, to determine the pareto optimal front. Once the pareto front is determined a few test points, along the pareto optimal front, can be used to determine the actual full simulation model objective function values, to scale the pareto optimal front.

It is suggested that a more effective approach for the determination of the pareto optimal front, would be to run the Dynamic-Q optimisation using a weighted objective function consisting of 100% handling and 0% ride comfort. Once the optimisation has converged to this optimum point, the algorithm should then change the objective function weighting to 80% handling and 20% ride comfort. By continually changing the objective function weighting

Table 9.1: Summary of Results for Optimisation Objectives

variables, opt. run	Fig.	# iter. (eq evals)	$f^*(\mathbf{x})$ ± 0.01	$\dot{\varphi}_{RMS}$ $[^o/s]$	φ_{peak} $[^o]$	a_{RMS_d} $[m/s^2]$	a_{RMS_p} $[m/s^2]$
Combined							
2, handling 1^{st}	9.1	6 (9.8)	0.21	1.29	2.9	-	-
2, ride after	9.1	18 (34.2)	0.40	1.52	3.0	2.20	2.18
2, $f_{max}(x)$	9.2	6 (12.6)	0.32	1.86	4.0	1.78	1.78
4, $f_{max}(x)$	9.3	7 (20.8)	0.32	1.83	4.0	1.76	1.62

Table 9.2: Summary of optimum damper factors and gas volumes

opt. run	Fig.	$dpsff$	$gvolf$	$dpsfr$	$gvolr$
Combined					
2, handling 1^{st}	9.1	1.35	0.10	1.35	0.10
2, ride after	9.1	0.55	0.17	0.55	0.17
2, $f_{max}(x)$	9.2	0.51	0.30	0.51	0.30
4, $f_{max}(x)$	9.3	0.53	0.26	0.40	0.28

during the optimisation procedure, the optimisation should progress along the pareto optimal front, from the best handling objective function value to the best ride comfort objective function value.

9.5 Conclusions

This chapter shows that the use of simplified numerical models, originally used for the optimisation of ride comfort and handling separately for gradient information, Chapters 5 and 6, can be successfully used for the combined optimisation of the ride comfort vs. handling compromise suspension configuration.

The advantages and disadvantages of different definitions of the objective

Table 9.3: ADAMS data for 4 test points along pareto optimal front

test point	$dpsff$	$gvolf$	$dpsfr$	$gvolr$	φ_{peak}	$\dot{\varphi}_{RMS}$	a_{RMS_d}	a_{RMS_p}
1	0.91	0.10	0.39	0.15	1.80	2.93	2.55	2.27
2	0.80	0.10	0.39	0.27	2.18	3.24	2.38	1.89
3	0.54	0.18	0.39	0.28	2.48	3.65	1.89	1.71
4	0.39	0.30	0.39	0.38	2.95	4.44	1.46	1.47

and constraint functions for the determination of the optimal suspension characteristics for combined ride comfort and handling are highlighted in terms of achievable optimal ride comfort and handling. It is, however, found that the use of only the simplified models to optimise for the pareto optimal front is most efficient. The design variable values along the pareto front can be used to determine the actual full simulation model's objective function values. The simplified model's pareto front is accurate in terms of design variable values, but the objective function values differ in absolute value.

It is suggested that a future implementation of continuously varying the weighting attached to the different objectives within the objective function definition will more effectively define the pareto optimal front. The optimisation convergence history, will then converge from the best of the one objective gradually towards the best of the other objective, along the pareto optimal front.

The methodology proposed is shown to be an efficient means of optimising a vehicle's suspension system for combined ride comfort and handling. This research illustrated that the use of gradient-based optimisation algorithms are suitable and competitive for determination of the pareto optimal front necessary for optimising vehicle suspension systems when considering combined ride comfort and handling.

Chapter 10

Conclusions

The use of central finite differences with relatively large perturbation sizes has proven to be beneficial in terms of total function evaluations needed to obtain a feasible minimum. This was measured in terms of less noise in the optimisation convergence history, although at an increased number of function evaluations per iteration, but less overall iterations to reach a feasible optimum. This approach holds definite benefit for all gradient-based optimisation algorithms.

A highly nonlinear vehicle model, with large suspension deflection, that returns excellent correlation to measured results was built in MSC.ADAMS. A novel lateral driver model that makes use of the magic formula to define a nonlinear steering gain factor, was proposed and successfully implemented. This nonlinear steering gain factor modelled with the magic formula made it possible to achieve excellent correlation with measured test data, for a single preview point yaw rate steering driver model. This driver model proved to be robust for different suspension setups, when optimising the vehicle's handling for the closed loop double lane change manœuvre.

The necessity of including wheel hop in the ride comfort optimisation problem was investigated. It was found that it is necessary to include wheel hop as an inequality constraint, when optimising the vehicle's suspension for ride comfort, if the vehicle is to remain stable on rough terrain.

Nonlinear simple models that capture the essence of the handling and ride comfort have been developed, and shown to exhibit similar trends to the full computationally expensive numerical simulation model of the off-road vehicle, at approximately 10 % of the simulation cost. The two design variable case of the simplified models had to be scaled to be equivalent to the full simulation model over the design space. These simplified models have been successful in speeding up the optimisation process, by at least 50% of total simulation time needed when using only the full simulation model for gradient, objective and constraint function values, when used for the determination of gradient information by means of central finite differences. This is a novel approach to vehicle suspension design optimisation and has been shown to be accurate and economical when compared to full simulation gradient based optimisation. The contribution in the field of vehicle design is also underlined by the fact that the same principle can be applied to any gradient-based optimisation algorithm.

The optimisation problem was expanded from 4 design variables to 14. Difficulties were encountered with poor scaling of the design variables, and noise associated with infeasible tyre deflections due to the current tyre model only accommodating a linear vertical tyre stiffness. Scaling of the optimisation problem has been investigated, with the result that sphericality of the design space is more important than having equivalent ranges and magnitudes of design variables. Great improvements were achieved in the optimisation convergence histories, when the optimisation problem was better scaled.

The scaling process that was followed, is reformulated into a novel automatic scaling methodology, that can help engineers reduce the time necessary for investigation of design variable scaling. This methodology was tested on analytic functions, and found to improve the optimisation convergence for most tested problems. The methodology was expanded to include constrained optimisation problems in the form of the penalty function, but further

experimentation is required with the penalty function multiplication factor to be used. The automatic scaling methodology was applied to the vehicle suspension optimisation for 14 design variables, with great success. When automatic scaling was applied, better optimum values were reached, than without automatic scaling.

There is at this stage no guarantees that the results achieved are global optima, as concluded from the fact that some of the optimisation results return similar objective function values for different design variable combinations. The aim is however, for an improvement in handling and ride comfort rather than the absolute global optimum suspension setup. There is also no guarantee that if the global optimum is found, the design variables are robust in terms of manufacturing tolerances.

Noise in the inequality constraints, when optimising ride comfort, and combined optimisation of ride comfort and handling, is still problematic. Some avenues were investigated but a more intensive investigation is needed before the problem is fully understood.

The combined optimisation of ride comfort and handling was investigated. Various concepts were investigated for the definition of the objective function. The discontinuous nature of the maximum function in the definition of the objective function was found to pose no difficulties in terms of optimisation convergence. When optimisation was performed using the baseline vehicle's handling as a constraint no improvement was found in ride comfort, and the same applies when the baseline vehicle's ride comfort was used as a constraint and handling optimised no improvement was observed. This was because the baseline vehicle's design point lay outside the feasible design space achievable with the current $4S_4$ suspension system when optimised for the compromise. This will probably be overcome when the control system is included in the optimisation process.

Chapter 11

Discussion of Future Work

The results of the 14 design variable optimisation, postulated that a solver change has the potential to greatly reduce the numerical noise present in the objective functions with respect to small perturbations of the design variables. This would be implemented by interpolation of the equivalent function value, for a constant numerical error, at each time step of the multi-body dynamics solver. It is believed that this alone will greatly contribute to reduced noise in the objective functions obtained from numerical simulations.

The vertical tyre stiffness should be modelled as nonlinear, so as to capture the effect of increasing tyre stiffness with high tyre deflections. This will be beneficial when considering the tyre hop inequality constraint at low suspension damping, which is currently resulting in optimisation convergence difficulties when considering many design variables. This, however, is not easily implementable in the current ADAMS Pacejka '89 tyre model used. Other tyre models would have to be investigated for the implementation of this stiffness characteristic.

A preliminary investigation has been performed in the use of gradient only optimisation algorithms like LFOPC (Snyman 2000) and ETOPC (Snyman 2005a), used with the simplified models, not presented in this thesis. This should be further investigated, as it could prove to be more efficient.

A greater variety of road conditions need to be considered over varying vehicle speeds and loading conditions, before a decision can be made regarding the final overall optimum design. The ultimate test will be the optimisation of the vehicle's performance under severe handling manœuvres on an uneven road. The methodology presented in this research is easily adaptable to these multiple conditions.

The incorporation of the complex model describing the hydro-pneumatic suspension's characteristics as proposed by Theron and Els (2005), should be included with the control of switching proposed by Els (2006), in the final optimisation phase.

The final optimised spring and damper characteristics should be investigated for robustness. This should be done in terms of the effect normal manufacturing tolerances will have on the vehicle's handling and ride comfort.

The proposed automatic scaling methodology should be further researched so as to take Hessian information into account before scaling the design space, in an effort to minimize the negative effects cross-coupled design variables have on the current scaling method.

A variable weighting when performing multi-objective optimisation needs to be investigated, so as to more efficiently plot the pareto optimal front, from the optimum of the one objective to the optimum of the other objective.