Improving Optimization-based Inverse Analysis using Direct Inverse Maps: A Dynamic Damage Identification Case Study

Grant Stephens and Daniel Nicolas Wilke

Emerging Engineering Technology Laboratory, Department of Mechanical and Aeronautical Engineering, University of Pretoria, Pretoria, South Africa, 0086; grant@stephens.co.za; nico.wilke@up.ac.za

Abstract: Inverse problems in engineering form routinely part of larger engineering simulations. Therefore, the quality of the solution to an inverse problem directly influences the quality of the larger simulation and, ultimately, the ability to solve an engineering problem. Inverse problems can be challenging and time-consuming to solve, as most inverse strategies require iteration due to the non-linear nature of the problem. As a result, they often remain poorly solved before proceeding to the larger analysis. The quality of the solution to an inverse problem is influenced by the inverse strategy, scaling of the problem, scaling of the data, and initial guesses employed for iterative strategies. Research has focussed considerably on inverse strategies and scaling. However, research into strategies that improve initial guesses of an inverse problem has been largely neglected. This study proposes an elegant strategy to improve the initial guesses for conventional optimization-based inverse strategies, namely direct inverse maps (DIMs) or inverse regression. DIMs form part of modern multivariate statistics. DIM approximates the solution to an inverse problem using regression; popular choices are linear regression, e.g., partial least squares regression (PLSR). These strategies are not iterative but require several independent a-priori simulations to have been conducted. As they are not iterative, one way to improve the solution is to increase the number of independent a-priori simulations to be conducted. Our proposed strategy is to use DIM to generate initial guesses for optimization-based inverse strategies. We conduct a parameter investigation on a truss structure's virtual vibration-based damage identification problem.

Keywords: Inverse Problem; Virtual Inverse Problem; Direct Inverse Maps; Partial Least Squares Regression; Optimization; Starting Point; Initial Guess

1 Introduction

Inverse analysis is prevalent in engineering analysis, as the characterization of models is routinely part of a larger analysis or simulation. The characterization of the models significantly influences the simulation quality. It is common for

engineers to focus on the analysis or simulation without properly characterizing some of the models used in simulations.

Inverse problems in engineering form routinely part of larger engineering simulations. Therefore, the quality of the solution to an inverse problem directly influences the quality of the larger simulation and, ultimately, the ability to solve an engineering problem. Inverse problems can be challenging and time-consuming to solve, as most inverse strategies require iteration due to the non-linear nature of the problem. As a result, they often remain poorly solved before proceeding to a larger analysis. The implication is usually disastrous, as conclusions drawn from the numerical work may not be valid. In addition, the solution quality for an inverse problem can be influenced by the inverse strategy, loss surface complexity, scaling of the data, and initial guesses employed. Most research has focused on inverse strategies and scaling of data. However, research into strategies that improve initial guesses of an inverse problem has been largely neglected.

This study focuses on and proposes an elegant strategy to improve the initial guesses for the ubiquitous optimization-based inverse strategies using the lesser-known direct inverse maps or direct inverse regression strategies [1]. DIMs can be broadly categorized as either iterative optimization-based or non-iterative regression-based approaches.

Optimization-based inverse strategies are well-known and ubiquitous in research and industry; in particular, weighted least-squares are considered a classical inverse analysis approach. Optimization-based inverse strategies start from an initial guess and iteratively improve the model parameters by minimizing some non-linear error and quantifying the difference between experimental and simulated responses until convergence. The sum of the error squared is a classical error measure often employed. Optimization-based inverse solution strategies can become computationally expensive when multiple local minima exist. Confidence in the solution is usually ensured by conducting multiple minimizations using gradient-based and evolutionary strategies in a multi-start procedure. In extreme cases, a multi-start approach may be required to obtain a converged or feasible solution, for instance, when an optimizer traverses model parameters that fail to analyze along a search path. However, a significant benefit of the minimization-based inverse strategies is that the error between the predicted and experimental response is reduced iteratively.

DIMs, in turn, are not so well known as they are rooted in modern multi-variate statistics [1, 2]. These strategies regress the experimental response directly to the model parameters without the need to iterate. Although these strategies are not iterative, they require a regression set of model parameters and their respective responses, usually constructed by simulation. DIM is, therefore, susceptible to extrapolation from the regression set when noise (aleatoric) and model (epistemic) mismatches exist between the experimental response and simulated responses in the regression set. The simulated response can then be computed using the model

parameters. Supplementing the regression set with additional simulated responses may improve the regression quality, but it is not guaranteed. The improvement depends on the similarity between the simulated responses in the regression set and the experimental responses regressed to the model parameters. The benefits and pitfalls of DIMs are detailed in [1]. Mature strategies quantify and address both aleatoric (noise) and epistemic (model mismatches) errors.

They include principal component regression (PCR) [3-5] and partial least squares regression (PLSR) [3, 6-8]. The optimization-based iterative and regression-based non-iterative inverse strategies are utilized in isolation [2, 9, 10]. The classical optimization-based inverse approach usually uses uniform or normally distributed initial starting points over an anticipated model parameter space [11]. Alternatively, judgment, experience, or exploiting physics can inform suitable starting guesses, but this requires focused and qualified analysts. Alternatively, DIMs merely regress the response to the model parameters using a regression set [2, 9, 12]. Essentially, when these strategies are used in isolation, we have

1) multiple minimizations that are conducted in isolation and

2) large sets of independent simulated responses are generated to make up the regression set.

This study proposes a complementary unified approach from these two approaches. We use DIM to predict starting points for a gradient-based inverse procedure that iteratively minimizes some nonlinear scalar error measure, which in this study is demonstrated for the sum of the error squared.

The benefit is that the computational cost to solve an inverse problem may be reduced and the robustness enhanced. We specifically consider partial least squares regression (PLSR) to predict the initial starting point when minimizing the sum of errors squared using a gradient-based minimizer. The robustness and the computational cost of solving optimization-based inverse problems are directly related to the quality of the initial starting points supplied to the optimizer. Of course, numerous strategies exist to improve the robustness of solving inverse problems. Examples include response surfaces [23] and lower fidelity models [21], which will benefit from improved starting points for the minimization strategy.

We consider a virtual inverse problem [21-22, 1] where we simulate the problem with and without simulated noise (aleatoric uncertainty) on the computed response. The benefit of knowing the solution is that we can properly quantify the performance of the various strategies in identifying the correct parameters. Since the considered approaches see the same problem, we can draw sensible conclusions in our comparisons between approaches. This study evaluates the dynamic identification of damage in a 25-bar truss structure [13].

2 Minimisation Inverse Strategy

Minimization-based inverse strategies predict the unknown model parameters by minimizing some scalar error measures between the experimental and simulated response.

The difference between the experimental and predicted response results in an error vector. The error vector can be reduced into a scalar form, e.g., the sum of the errors squared or stated as the square of the L_2-norm, resulting in the familiar least squares problem. The least squares problem is an unconstrained minimization problem usually solved using a Quasi-Newton minimizer [23].

Alternatively, the error vector can be the constraint in a min-max optimization problem where the maximum error is minimized. Here, the L_∞ norm is usually minimized [15]. The objective function is linear by construction, but each point in the predicted response is a constraint. This constrained problem is generally solved using an augmented Lagrangian strategy [23].

This study is limited to least squares minimization but can be applied to various error measures.

3 Direct Inverse Maps

DIM regresses an observed experimental response to unknown model parameters. Various DIM strategies exist. The most popular are based on high dimensional linear regression constructed from a regression set. The regression set contains model parameters and their respective responses. The regression set is usually computed by simulation of a representative model. The regression strategies include principal component regression (PCR) [3-5, 8] and partial least square regression (PLSR) [3, 6-8].

These approaches are not susceptible to co-linearity problems like multi-linear regression (MLR) as they also reduce the dimension of the problem by projection. Hence, many high-dimensional points can be used in the regression set [24]. Both PCR and PLSR require the size of the reduced dimension (number of modes or loadings) to be selected by the user [24]. For regression, it is often suggested to use the lowest number of modes such that the responses in the regression set describe the experimental response well. This implies that the projected experimental response onto the modes can be expressed by interpolating between the responses in the regression set that are also projected onto the modes. This is opposed to extrapolation, which indicates the response is an outlier to the points in the regression set.

In this study, we only consider partial least square regression. PLSR solves the following problem: and is given by the following algorithm PLSR is available in most numerical software packages such as Matlab (function `plsregress` in the Statistics and machine learning toolbox), R (`pls` package), and Python (module `PartialLeastSquares`).

4 Combined Strategy

Optimization inverse analysis and DIMs are generally used in isolation due to their historical origins. Instead of comparing these two approaches for their benefits and pitfalls, we demonstrate the potential benefits when unifying them to solve inverse problems. This approach can be applied when the simulation model is evaluated directly [22] or via a response surface [24]. In this study, we only consider the former approach to allow us to assess its merit without introducing additional complexities and uncertainties, such as the quality of the response surface. We note that should this approach prove beneficial in the absence of response surfaces, the additional benefits are evident when including response surfaces [23 - 25]. The simulated data points to construct a response surface can also generate initial starting points that might be less susceptible to local minima and more likely to lie within the global basin [26].

Our proposed strategy under investigation uses the PLSR as a DIM to compute the initial starting point for a classical Quasi-Newton gradient-based minimization algorithm. This process is repeated until convergence. Consider the algorithmic outline of our approach:

1 Estimate the expected domain for each input parameter $\boldsymbol{X}$ of the model. Initialize an empty design of experiments (DOE) $\boldsymbol{XY}$. Choose N_{DOE} and N_{OPT}.

2 Compute N_{DOE} points and augment the DOE $\boldsymbol{XY}$, using augmented Latin Hypercube Sampling (ALHS) over the expected parameter domain.

3 Predict the model input parameters $\boldsymbol{X}_P$, from the experimental response $\boldsymbol{Y}_{EXP}$ using PLSR.

4 The predicted response is then used as an initial starting point for an iterative minimization approach x0 = XP when there is a lower error than the best point in the regression set. Otherwise, the point with the lowest error in the regression set is used as an initial starting point. The minimizer is limited to N_{OPT} iterations. Here, we can include or discard the responses computed at each iteration to be used in the regression set for the next PLSR. In this study, we only include the point at iteration N_{OPT}.

5 Repeat steps 2-5 until convergence.

Note that if we choose $N_{DOE} = 1$, we recover an optimization-based inverse strategy with a single initial random guess. By choosing $N_{OPT} = 0$, we recover PLSR solely. This allows us to quantify the benefits of the independent and various blends of the unified strategy.

An additional parameter to $\underline{N_{DOE}}$, $N_{OPT,}$ and the convergence criteria is the number of modes N_{MODES} to use for the PLSR. The number of modes N_{MODES} are estimated as follows:

1 Compute the participation coefficients (scores) of each mode (loading) for each point in the regression set.

2 For each loading estimate, the expected ranges for the participation coefficients.

3 Project the experimental response $\boldsymbol{Y}_{EXP}$ onto the first mode using standard least squares regression and compute the participation coefficient.

   a. If the participation coefficient of the first mode falls outside the expected range, use the data point in the set used to construct the PLSR with the least error as an initial optimization starting point; otherwise, continue.

4 Include the next mode in the projection and compute the participation coefficients.

5 Conduct the following convergence checks

   a. Check if the participation coefficients fall within the expected ranges.

   b. Check that the norm of the difference between the current and previously predicted model parameters decreased.

   c. Continue until a check fails and use all modes, excluding the last mode that failed a check.

6 Repeat 1-5 until a convergence check fails.

5 Numerical Investigation

We conduct a numerical investigation to assert the feasibility of solving inverse problems using the proposed approach of combining DIMs and minimizing the sum of the errors squared. By appropriately selecting N_{DOE} and N_{OPT}, we can investigate the two strategies in isolation and combined. In this study, we restrict ourselves to a virtual inverse problem [21-22, 1], using simulated experimental data instead of actual experimental data. The benefit is that we can critically assess the quality of the estimated model parameters and the responses' quality.

We effectively turn an unsupervised learning problem into a supervised learning problem through modeling.

We conduct the inverse analysis with simulation in the loop instead of the typical and appropriate response surface-based strategy [25-26]. This is done deliberately to investigate the proposed approach without adding the additional complexity and uncertainty that response surfaces may introduce into the investigation.

We consider the vibration-based damage identification of a 25-bar truss. We aim to estimate the mass of 25 trusses from the first two vibration modes given by the nodal displacements of the six nodes of the truss structure, resulting in a 36-dimensional response vector. We consider the simulated experimental response vector with and without simulated measurement noise.

5.1 Vibration-based Damage Identification of a 25-bar Truss

Vibration-based damage identification approaches aim to assess the integrity of structures non-destructively [13, 16-18]. Modal parameters (natural frequencies, mode shapes, damping, and modal strain energy) of a structure depend on the physical properties of the structure, i.e., mass, stiffness, and damping. The premise behind vibration-based damage identification is that changes in the physical properties manifest in changes to the modal parameters. Hence, changes in modal parameters can be used to locate and identify damage in a structure.

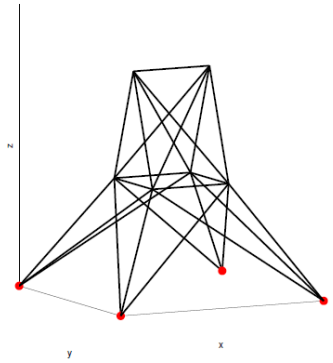
We consider a simple example in which we aim to estimate the effective mass (or equivalently, area) of the truss members from only the first two structural modes of a structure in a corrosive environment. The geometry and material properties are assumed to be known. A truss analysis without damping is considered sufficient to describe the dynamics of the structure, i.e., a lightly damped structure. The nodal displacements of the six nodes in three-dimensional space for two modes result in a 36-dimensional response vector.

The 25-bar truss structure [19] is depicted in Figure 1(a), and the first two modes are in Figures 1(b) and (c), respectively. The structure is fully constrained (in all three translational directions) at the four ground supports, as indicated in Figure 1(a). The truss modal analysis is conducted using an in-house finite element analysis code with direct and adjoint analytical sensitivities developed by Wilke for his Optimum Design (MOO780) graduate course at the University of Pretoria, South Africa, which solves the generalized eigenvalue problem

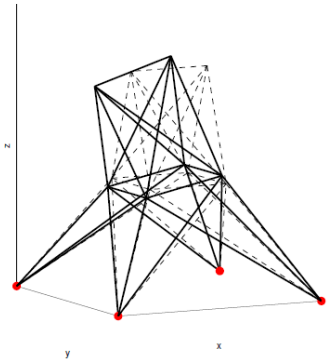
$$(\boldsymbol{K}_{ff} - \lambda \boldsymbol{M}_{ff})\boldsymbol{y} = \boldsymbol{0}, \tag{1}$$

with $\boldsymbol{K}_{ff}$ and $\boldsymbol{M}_{ff,}$ the assembled system stiffness and mass matrices with the associated unconstrained degrees of freedom of the system. The experimental data in this vibration-based damage identification problem is simulated using the same code to conduct a virtual inverse problem.

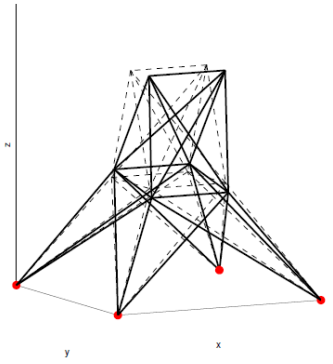
This study considers simulated experimental data with and without measurement noise, i.e., aleatoric uncertainty. A normally distributed measurement noise of 2% for the first standard deviation was assumed. The experimental data is simulated for 1 kg masses for each of the 25 truss members, which results in a 487 Hz fundamental frequency and a second mode at 530 Hz. An unknown parameter range of 100% around the solution is considered for this study. The parameters to be estimated were normalized by the solution [20]. The error norm for the calculated response is 0.0 (within machine precision) in the absence of noise and around 0.5102 when 2% measurement noise is assumed.



(a) Undeformed truss structure with fixed points denoted by red dots.



(b) First (fundamental) vibration mode of the truss structure (487 Hz).



(c) Second vibration mode of the truss structure (530 Hz).

Figure 1

(a) The undeformed truss structure and associated (b) fundamental at 487 Hz and (c) second mode at 530 Hz. The base (indicated by the solid red dots) is fixed to resist any translation.

6 Numerical Results

The numerical investigation aims to quantify our proposed strategy's performance and sensitivity to divide the effort between adding points to the N_{DOE} to improve starting points instead of allowing additional optimization iterations N_{OPT}. We investigate the benefits of splitting effort between the PLSR and iterative optimization without explicitly considering the associated computational cost to avoid distracting from this study. We limit ourselves to 500 function evaluations and consider the cost of each optimization iteration to be one function evaluation. Similarly, each point added to the DOE is regarded as one function evaluation. The benefit of this choice is that a user can scale the associated computational cost of an optimization iteration to DOE point computation independently afterward.

We consider a purely sequential computational framework, but the benefits of parallel environments are evident when increasing the number of points of the DIM, as it is embarrassingly parallel as opposed to the sequential nature of a classical gradient-based algorithm. However, these questions warrant an independent study to explore when a multi-core, GPU, or parallel computational architecture is considered. The implicit assumption of our choice is that function computations dominate the computational cost of these two problems, with sensitivities being available computationally efficiently either analytically or by differentiating a response surface representation of the cost function. In this study, adjoint analytical sensitivities are computed.

To conduct our study, we consider distinct ratios $N_{OPT}:N_{DOE}$ of the number of optimization iterations N_{OPT} to Latin hypercube sampled N_{DOE} points. The ratios we consider for both problems are 0:9, 1:9, 4:9, 1:1, 9:4, 9:1, and 100:1. A ratio of 0:9 implies that no optimization is conducted, and the response is computed using only PLSR. In turn, a ratio of 100:1 effectively results in an optimization strategy using random initial starting points. We conduct this for increments of $\Delta N_f = \{10, 25, 50, 100, 250, 500\}$ function evaluations per iteration for our combined strategy. When only PLSR is considered, we compute exactly ΔN_f function evaluations, whereas the optimization strategy with random initial starting points is limited to a maximum number of ΔN_f function evaluations. This is repeated until the maximum of 500 function evaluations is reached. Table 1 indicates the number of function evaluations available for PLSR and the least squares minimization strategy for the different choices of $N_{OPT}:N_{DOE}$ and ΔN_f.

The optimization algorithm used in this study is Matlab's SQP algorithm in the `fmincon` function. The convergence tolerance for the optimization algorithm was set to 10^{-6} for changes in updates or function values between updates.

Table 1
The available number of function evaluations (N_{DOE},N_{OPT}) for the combined strategy per iteration for the selected ratios N_{OPT}:N_{DOE} and ΔN_f

		ΔN_f					
		10	25	50	100	200	500
$\frac{N_{OPT}}{N_{DOE}}$	0	(10,0)	(25,0)	(50,0)	(100,0)	(200,0)	(500,0)
	1 : 9	(9,1)	(22,3)	(45,5)	(90,10)	(180,20)	(450,50)
	4 : 9	(7,3)	(17,8)	(35,15)	(69,31)	(139,61)	(346,154)
	1 : 1	(5,5)	(12,13)	(25,25)	(50,50)	(100,100)	(250,250)
	9 : 4	(3,7)	(8,17)	(15,35)	(31,69)	(61,139)	(154,346)
	9 : 1	(1,9)	(3,22)	(5,45)	(10,90)	(20,180)	(50,450)
	100 : 1	(0,10)	(0,25)	(0,50)	(1,99)	(2,198)	(5,495)

For each selected setting, we repeat the optimization run 100 times and depict the results in box plots for choices of ΔN_f. The box plots show the median (solid red line), with the box representing half the data points. The whiskers extend to the extreme data points that are statistically unimportant. Lastly, the red crosses indicate all the outliers.
We quantify the analyses' robustness and solutions' quality. The analyses of the 25-bar truss always succeeded.

6.1 Vibration-based Damage Identification of a 25-bar Truss

The final residuals obtained for the 100 independent runs after 500 total function evaluations are presented in Figures 2(a)-(g) and 3(a)-(g) for the noiseless and noisy simulated experimental data, respectively. Figure 4 shows the results after only 100 function evaluations to highlight the relative improvement from 100 to 500 function evaluations. In Figure 2(a), the PLSR predictions vary between two error levels. It is evident from the median position that divides the data points at the top or bottom of the box. In Figure 2(a), as expected, the same response for all ΔN_f is evident when considering the size and position of each box.

There is a definite benefit to selecting a lower NOPT:NDOE ratio for the noiseless simulated experimental data, except for a ratio of 0, which represents only PLSR. A larger ΔN_f seems to be consistently beneficial. The 0 ratio effectively represents the contribution of the PLSR in isolation, whereas the minimizer effectively realizes the remainder of the improvements. Note that a single optimization run is preferred, as consistently indicated by the better performance of the larger ΔN_f. It demonstrates the benefit of continuing the minimization after an appropriate initial guess has been provided. Additional PLSR regression points perform worse

than allowing additional iterations in the minimizer. However, an estimate of a proper initial point is important as minimization in isolation, similar to PLSR in isolation, performed the worst on average.

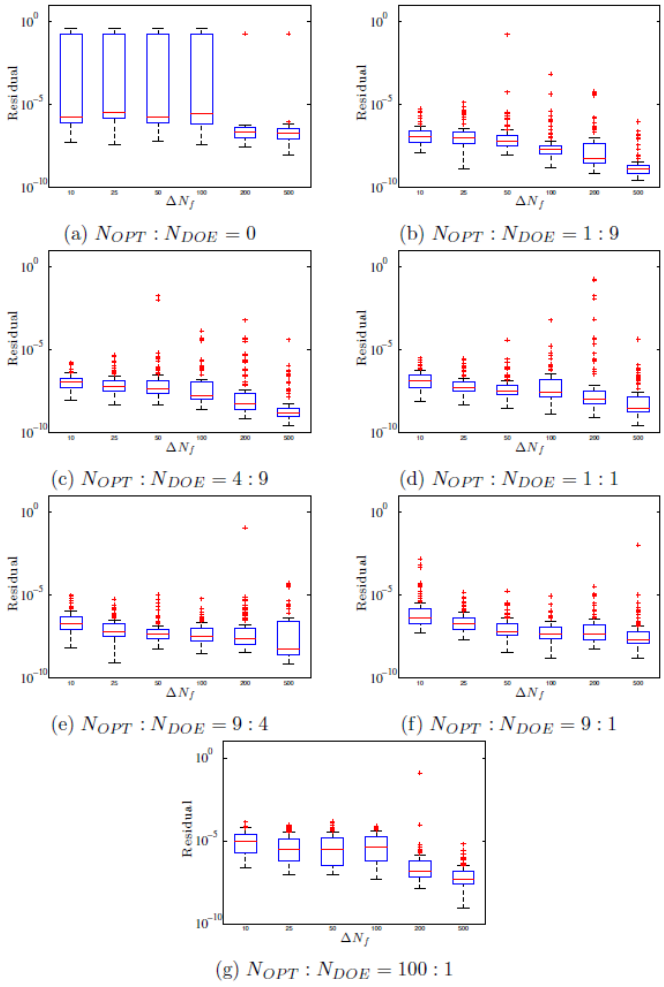


Figure 2

Residual box plots for 100 independent runs using the simulated experimental data without noise for the damage identification problem. The total number of function evaluations is 500 for the selected ratios N_{OPT}:N_{ODE} (a)-(g). The choice ΔN_f is quantified in each subfigure.

This contrasts the simulated experimental data with noise, where the selection of a moderate ΔN_f yields better results on average. In addition, a ratio N_{OPT}:N_{DOE} that favors the minimizer is beneficial. Note that suitable ratios N_{OPT}:N_{DOE} allow for PLSR-estimated starting points instead of random initial starting points.

The implication is that multi-starts benefit this noisy problem, as indicated by a low to moderate value for ΔN_f.

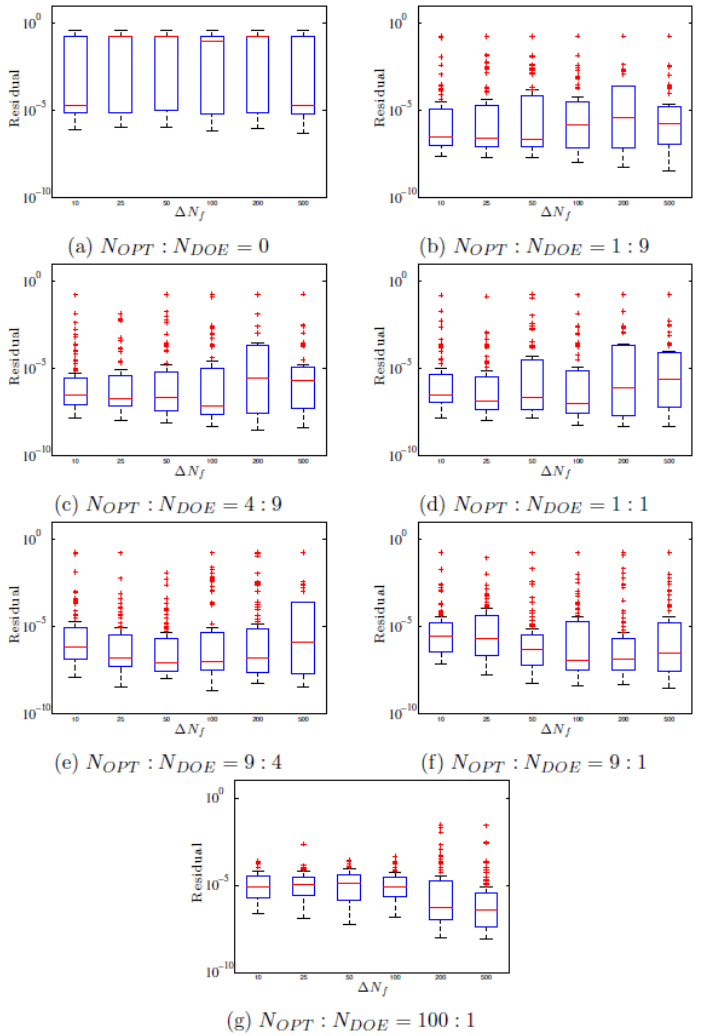


Figure 3

Residual box plots for 100 independent runs using the simulated experimental data with noise for the damage identification problem. The total number of function evaluations is 500 for the selected ratios N_{OPT}:N_{ODE} (a)-(g). The choice ΔN_f is quantified in each subfigure.

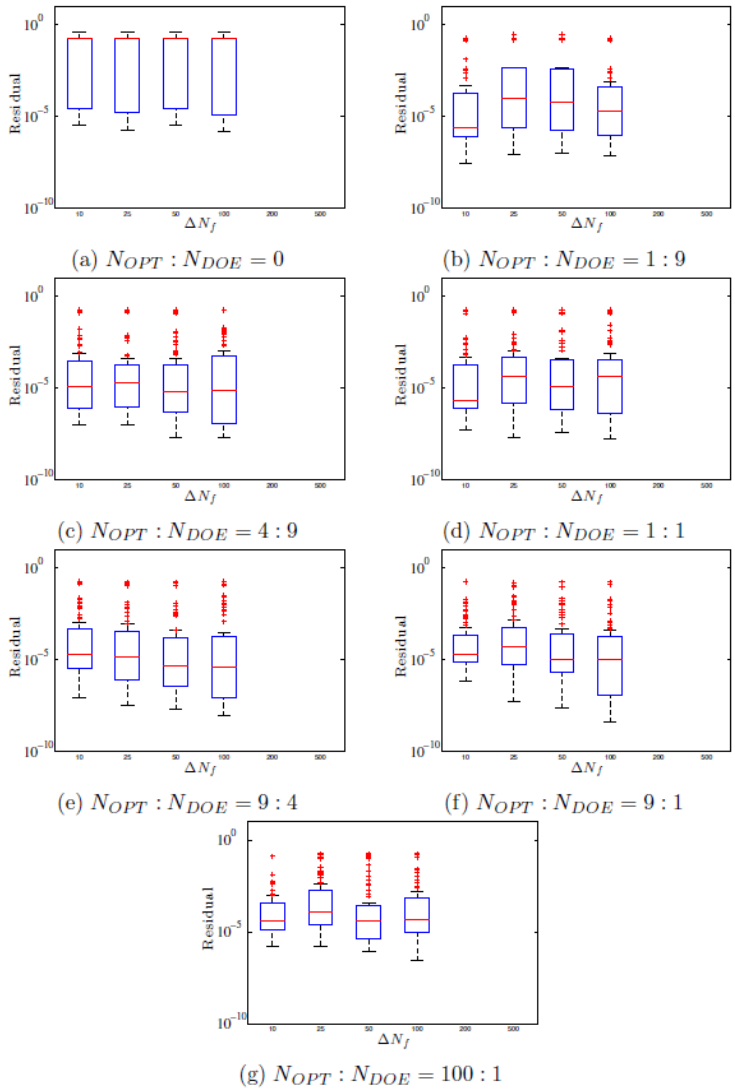


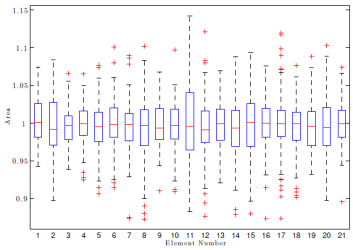
Figure 4

Residual box plots for 100 independent runs using the simulated experimental data with noise for the damage identification problem. The total number of function evaluations is 100 for the selected ratios N_{OPT}:N_{ODE} (a)-(g). The choice ΔN_f is quantified in each subfigure.

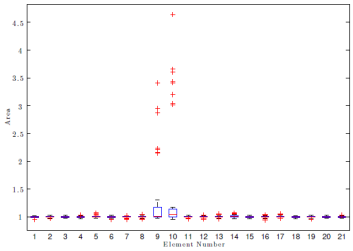
Minimization in isolation (N_{OPT}:N_{DOE} = 100:1) and PSLR in isolation (N_{OPT}:N_{DOE} = 0) performs the worst. Notably, in isolation, the minimizer significantly reduces the number of outliers at the cost of a higher residual on average. The results obtained after 100 total function evaluations for the noisy simulated experimental

data are presented in Figure 4(a)-(g). This shows the progression of the decrease in residual and also how the various settings benefit from an additional 400 function evaluations that Figure 3(a)-(g) depicts.

We also present area variation as box plots in Figures 5(a) and 5(b) for a range of residual values. The solution is obtained when all the areas are 1, as seen in Figure 5(a). All values, including outliers, are within a 15% range of the solution. From Figure 5(b), the larger residuals are due to the minimizer poorly estimating the truss areas and not the PLSR-estimated initial guesses.



(a) Boxplot of Areas for each of the elements found by $N_{OPT} : N_{DOE}$ =1:9 and $\Delta N_f = 25$ with a range of residuals between 1.33×10^{-5} and 1.24×10^{-9}



(b) Box-plot of Areas for each of the elements found by $N_{OPT} : N_{DOE}$ =1:1 and $\Delta N_f = 200$ with a range of residuals between 0.1780 and 8.05×10^{-10}

Figure 5
Box-plots of the areas found by two different sets of data

To investigate the quality of the computed initial guesses as a function of the number of LHC points, we compute 100 initial guesses for various numbers of points in the regression set. The results are depicted in Figure 6. Note the significant benefit of a small sample size in the regression set.

The results presented demonstrate a computational benefit when unifying optimization-based inverse strategies and DIM. A benefit that is enhanced in the presence of experimental measurement noise.

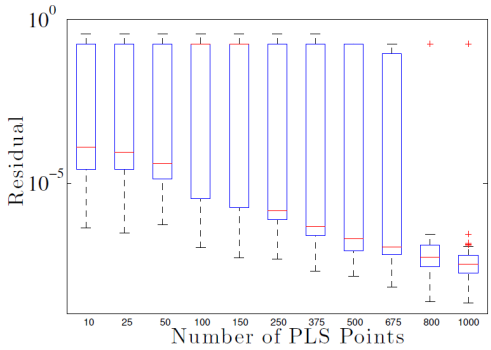


Figure 6
Boxplot of the residuals computed for 100 initial guesses for the optimization algorithm as the number of points in the regression set increases

Conclusion

This study demonstrated the benefit of unifying two inverse strategies: least squares minimization and the PLSR direct inverse map (DIM). Traditionally, these two strategies are considered separately, but this study demonstrates the benefits of unifying them into a complementary approach. Our proposed strategy unifies these two into a single strategy that allows each to be recovered in isolation. DIM is used to compute suitable initial guesses for the iterative optimization solver. For a practical inverse problem, we demonstrated that this combination results in computational benefits.

This unified approach demonstrated benefits when considering a sequential computing platform. The additional benefits of this approach on parallel computing platforms will only complement this strategy. This will be investigated in a future study to investigate and quantify each properly.

Future work will consider the benefits of this strategy in terms of robustness, as preliminary work on identifying ODEs on Calcium signaling pathways [14] only solves around 50% of the time when random starting points are used compared to 100% when combined with DIM. As future work will demonstrate, this unified approach can improve the robustness of the inverse problem when facing challenging problems or when limited resources are available.

References

[1] Erfan Asaadi, Daniel N. Wilke, P. Stephan Heyns, and Schalk Kok. The use of direct inverse maps to solve material identification problems: Pitfalls and solutions. Struct. Multidiscip. Optim., 55(2):613-632, feb 2017

[2] Wold S., Ruhe A., Wold H., and Dunn W. J. The collinearity problem in linear regression. the partial least squares (PLS) approach to generalized inverses. SIAM Journal on Scientific and Statistical Computing, 1984

[3] W. J. Dunn III, D. R. Scott, and W. G. Glen. Principal components analysis and partial least squares regression. Tetrahedron Computer Methodology, 2(6):349-376, 1989

[4] S. de Jong. SIMPLS: An Alternative Approach to Partial Least Squares Regression. Chemometrics and Intelligent Laboratory Systems, 18(6):251-263, 1989

[5] I. Jolliffe. Principal Component Analysis. Springer-Verlag, 2nd edition, 2002

[6] P. Geladi and B. R. Kowalski. Partial least-squares regression: a tutorial. Analytica Chimica Acta, 185:1-17, 1986

[7] R. Rosipal and N. Kramer. Subspace, Latent Structure and Feature Selection: Statistical and Optimization Perspectives Workshop (SLSFS 2005). Lecture Notes in Computer Science 3940. Springer-Verlag, 2006

[8] Hong-Dong Li, Yi-Zeng Liang, and Qing-Song Xu. Uncover the path from PCR to PLS via elastic component regression. Chemometrics and Intelligent Laboratory Systems, 104(2):341-346, 2010

[9] F. Abbassi, T. Belhadj, S. Mistou, and A. Zghal. Parameter identification of mechanical ductile damage using artificial neural networks in sheet metal forming. Materials & Design, 45:605-615, 2013

[10] M Rabinovich and D. M. Blei. The inverse regression topic model. In 31st International Conference on Machine Learning, 2014

[11] Abeeb a. Awotunde. Estimation of well test parameters using global optimization techniques. Journal of Petroleum Science and Engineering, 125:269-277, 2015

[12] Huber N. and Tsakmakis C. A neural network tool for identifying the material parameters of a finite deformation viscoplasticity model with static recovery. Computer Methods in Applied Mechanics and Engineering, 191:354-384, 2001

[13] D. Dessi and G. Camerlengo. Damage identification techniques via modal curvature analysis: Overview and comparison. Mechanical Systems and Signal Processing, 52-53, 2015

[14] C Solem, B Koebmann, and P R Jensen. Control analysis of the role of triosephosphate isomerase in glucose metabolism in Lactococcus lactis. IET systems biology, 2(2):64-72, 2008

[15] K. Madsen and H. Schjaer-Jacobsen. Algorithms for worst case tolerance optimization. IEEE Transactions of Circuits and Systems, CAS-26, 1979

[16] S. W. Doebling, C. R. Farrar, and M. B. Prime. A summary review of vibration-based damage identification methods. Structural Health Monitoring, 30, 1998

[17] W. Fan and P. Qiao. Vibration-based damage identification methods: a review and comparative study. Structural Health Monitoring, 10, 2011

[18] Y. M. Fu and L. Yu. A de-based algorithm for structural damage detection. Advanced Materials Research, 2014

[19] L. Schmit and C. Fleury. Discrete-continuous variable structural synthesis using dual methods. AIAA Journal, 18, 1980

[20] P. E. Gill, Murray W., and M. H. Wright. Practical optimization. Academic Press, 1981

[21] DN Wilke, S Kok, AA Groenwold. The application of gradient-only optimization methods for problems discretized using non-constant methods. Structural and Multidisciplinary Optimization, 40, 2010

[22] S Ben Turkia, DN Wilke, P Pizette, N Govender, NE Abriak. Benefits of virtual calibration for discrete element parameter estimation from bulk experiments. Granular Matter, 21, 2019

[23] JA Snyman, DN Wilke. Practical mathematical optimization: Basic Optimization Theory and Gradient-Based Algorithms, Springer Optimization and Its Applications (SOIA, Volume 133), Springer, 2018

[24] Y Chae, DN Wilke. Sub-dimensional surrogates to solve high-dimensional optimization problems in machine learning, Advances in Artificial Intelligence: Reviews, Volume 1, International Frequency Sensor Association Publishing, 2019

[25] D Correia, DN Wilke. How we solve the weights in our surrogate models matters, Journal of Mechanical Design 141 (7), 074501, 2019

[26] D Correia, DN Wilke. Purposeful cross-validation: a novel cross-validation strategy for improved surrogate optimizability, Engineering Optimization 53 (9), 1558-1573, 2021