

LEARNING PATTERNS OF EXPERT BEHAVIOUR IN
MULTI-OBJECTIVE DESIGN

A Thesis Submitted

in Partial Fulfilment of the Requirements

for the Degree of

Master of Technology (M.Tech)

by

Kushlendra Mishra

Supervised By

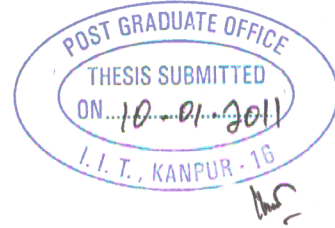
Dr. Amitabha Mukerjee



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY KANPUR

January, 2011

CERTIFICATE



It is certified that the work contained in this thesis entitled
“Learning Patterns of Expert Behaviour in Multi-Objective Design”,
by *Kushlendra Mishra (Roll No. Y8111018)*, has been carried out under my
supervision and that this work has not been submitted elsewhere for a degree.

(Dr. Amitabha Mukerjee)

Department of Computer Science and Engineering,
Indian Institute of Technology Kanpur

Kanpur-208016

January, 2011

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my thesis supervisor Dr. Amitabha Mukerjee for giving me an opportunity to work with him. His constant support, motivation, patience and insistence on exploring every aspect of the problem have been instrumental in bringing this thesis to a fruitful conclusion. I would like to take this opportunity to thank Dr. S. P. Das for enlightening discussion and valuable suggestions.

I am grateful to my parents for their faith, love and values that have always inspired me to walk upright. I would like to thank my brothers and sisters for their unconditional love and support. Words fall far too short of expressing what I want to for my Dadabhai, so I just thank him for being my hero, embodiment of inspiration, pillar of strength, guide ...

I would like to thank Ankit, Anil, Rahul, Ankur, Preejesh and all other friends for the long walks and endless non-technical discussions over innumerable cups of chai. I would also like to thank Gaurav and Hashmi for their friendship and support.

Finally, I would like to thank IIT Kanpur for providing excellent facilities and carefree learning environment.

Kushendra Mishra

Dedicated to
My parents and family.

Contents

Acknowledgements	iii
List of Figures	viii
List of Tables	xii
Abstract	xv
1 Introduction	1
1.1 Expert behaviour	1
1.1.1 A real design situation	2
1.2 Modelling the chunking behaviour	3
1.3 Chunk dimensionality conjecture	5
1.4 Dimensionality reduction	7
2 Proposed Methodology	10
2.1 Overview of the chunking process	10
2.2 Obtaining set of optimal designs	11
2.2.1 Conventional multi-objective optimization techniques	12
2.2.2 Evolutionary algorithms for multi-objective optimization	13
2.2.3 The BDCPM motor design problem	14

2.2.4	Local search procedure	18
2.3	Estimating the pareto-front manifold	19
2.3.1	Isomap Residual Variance for the Pareto-front	20
2.3.2	PCA explained variance	20
2.4	Clustering the pareto-front	21
2.5	Manifold modelling of the clusters	25
2.6	Design implications	28
2.7	How the chunks relate to experts' intuition	29
3	The Gearbox Design Problem	31
3.1	Problem description	31
3.1.1	NSGA-II formulation and the pareto-front	35
3.2	Fixed transmission ratio problem	36
3.2.1	Clustering the pareto-front	37
3.2.2	Discussion	39
3.3	29 Variable problem	42
3.3.1	Analysis of the clusters	43
3.3.2	Discussion	46
4	More Design Problems	48
4.1	Problem Introduction	48
4.1.1	Multiple-disk clutch brake design problem	48
4.1.2	The welded beam design problem	51
4.2	Analysis of the clutch brake design problem	52
4.2.1	Isomap and PCA analysis of the pareto-front	52

4.2.2	Clustering analysis	54
4.2.3	Discussion	56
4.3	Analysis of welded beam design problem	57
4.4	Isomap and PCA analysis of the pareto-front	58
4.4.1	Clustering analysis	59
4.4.2	Discussion	60
5	Conclusion	62
	Bibliography	65

List of Figures

2.1	Overview of the chunking process.	11
2.2	BDCPMM Stator assembly.	15
2.3	BDCPMM Rotor Assembly.	16
2.4	Pareto-front for the BDCPMM design problem. The red circles represent the design results of NSGA-II and the blue circles represent the design results of the local search. Pareto-front obtained after the local search overlaps with the NSGA-II result.	18
2.5	Isomap and PCA results. The residual variance plot shows the largest drop for the second dimension indicating a manifold of dimension one. First principal component has an 95% explained variance indicating a linear dimensionality of one.	21
2.6	3-dimensional Isomap Embedding of the BDCPMM problem pareto-front, showing several one-dimensional manifolds.	22
2.7	Clusters in the BDCPMM pareto-front. Clusters A and B are composed of low power and cost designs while Cluster D as the high-power and cost designs.	25

2.8	Clusters in the n_t - N subspace. All the clusters appear as lines in the parameter space except the cluster of high-end designs D which appears as two separate lines.	26
2.9	Residual variance for the cluster C . The largest drop in the residual variance is for the 3-dimensional Isomap embedding, indicating a two dimensional manifold.	27
3.1	18-speed Gearbox layout. Gears in shafts 1, 3 and 5 can be translated to mesh with corresponding gears in shafts 2, 4 and 6 respectively. With three different possible meshing of gears between shafts 1-2, 2-3 and two in shafts 2-4, 18 different output speeds are possible.	32
3.2	Isomap and PCA results for the fixed lay-out problem. The largest drop in residual variance is for the four dimensional Isomap embedding. The explained variance plot shows only one significant component.	37
3.3	Fixed layout gearbox design pareto-front and clusters in the objective space. There are 927 solutions in the pareto-front. The clusters in the extremes of the pareto-front are small.	38
3.4	Isomap and PCA results for the clusters of fixed lay-out problem. D has the largest residual variance, all other clusters have negligible residual variances. The plot indicates a one dimensional manifold. Only cluster D has two significant principal component, all others have only one significant principal component.	39
3.5	Power Vs. module characteristics of the pareto-front. Clusters occupy contiguous steps in the staircase structure.	41

3.6	The pareto-front and the clusters for the 29-variable problem. The pareto-front manifold is clearly a two dimensional manifold in the objective space.	42
3.7	Isomap and PCA results for the 29 variable gearbox design problem pareto-front. Although the largest drop in residual variance is for two-dimensional embedding, it keeps dropping up to the fifth dimension. PCA explained variance shows four significant principal components.	43
3.8	Isomap residual variances for clusters for the variable gear teeth problem. Residual variances for the clusters are similar to the whole pareto-front	44
3.9	Cumulative PCA explained variance for 29 variable problem. For most clusters, first five-six principal components cover 90% of the explained variance.	45
3.10	Power Vs. module characteristics of the pareto-front for 29 variable problem. Clusters show intermingling unlike for the fixed layout problem.	47
4.1	Schematics of a multiple disk clutch brake.	49
4.2	Schematics of a welded beam.	51
4.3	Pareto-front and the clusters of the multiple-disk clutch brake design problem. Clusters obtained with $k = 1.9$	53
4.4	Isomap and PCA results for the pareto-front of multiple-disk clutch brake problem pareto-front. The manifold dimensionality is one as indicated in the Isomap residual variance plot but the PCA explained variance shows two significant principal components.	54

4.5	Isomap and PCA results for the clusters of multiple-disk clutch brake problem. The manifold dimensionality of all the clusters is one, though cluster A has linear dimensionality of two indicated by two significant principal components in its explained variance plot.	55
4.6	Radius Vs. no. of friction surface plot for pareto-front clusters. Clusters A , B , C and F are distributed in four lines each while D , E and F are composed of single lines.	57
4.7	Pareto-front and clusters of the welded beam design problem. Obtained with $k = 2$. The clusters in the extremes are the largest while those around the knee of the curve are relatively smaller.	58
4.8	Isomap and PCA results for the welded beam design problem pareto-front.	59
4.9	Isomap and PCA results for the clusters of welded beam design problem. All the clusters are one dimensional manifolds but clusters C and D have two linear dimensions as they have two significant components. 60	

List of Tables

1.1	Summary of experiments.	7
2.1	Dimensions of the BDCPM motor family	15
2.2	First two principal components of the BDCPM data. Number of laminations is the dominant variable in the first principal component.	20
2.3	Explained variances for the principal components of the clusters. For clusters A , B , D and E there is only one significant principal component. Cluster D has two significant principal components.	27
2.4	Principal components for the clusters. For A B D and E the principal component is along the direction of n_l . Cluster C has two components which are in the $n_l - N$ plane.	27
3.1	No. of teeth for each gear-pair in fixed layout gear-train design.	35
3.2	First two principal components of the fixed layout gearbox pareto-front. t_9 and t_8 are the variables with the highest weights in the first component.	37

3.3	Principal component weights of each cluster in sorted order of absolute weights. Background shading indicates the input to output speeds for the gear pair. Darker cells indicate higher ratio. The gear pairs in the last transmission stage have the highest weights and those in the first transmission stage have the lowest weights for all the clusters.	40
3.4	Gear pairs with lowest input to output speed ratios.	40
3.5	Highest absolute weights of the first four principal components of the gear train design problem (29 variables). There is no dominant variable in the first principal component.	44
3.6	Ranking of gear teeth variables on the basis of absolute weights in principal components. The gear pairs of the last transmission stage occupy the last positions indicating they are the least varying. . . .	46
4.1	First two principal components of the clutch brake design problem pareto-front.	54
4.2	Significant principal components of the clutch brake design problem clusters. t and F have no weights in any principal component. The inner and outer radii(r_i and r_o) show the highest variation for each cluster. The second component is parallel to the <i>number of friction surfaces</i> (Z) dimension.	56
4.3	Principal component of the clutch brake design problem pareto-front	58

4.4 Significant principal components of the pareto-front clusters of the welded beam design problem. The one dimensional clusters have b and l as the predominant variables. Two dimensional clusters also have their significant components in the $b-l$ plane. t is the least varying variable in all the clusters. 61

ABSTRACT

For problems in their domain of expertise, human experts are known to arrive at a few good solutions immediately without exhaustive search. It is believed that such expertise depends on the learning of optimal chunks that emerge during the routine exploration of the problem space, constituting a dimensionality reducing representation for the task domain. Acquiring such expertise in a task domain is based on discovering functional interrelations among parameters in a decision space. In most situations, function is captured through multiple performance objectives, as a result of which no single optimum may exist.

In this work, we propose that the chunks arise as dimensionality-reducing structures that reflect clusters in the pareto-optimal region. We propose a *Chunk Dimensionality Conjecture*, by which for well-posed objective functions, the dimensionality of the chunk may be of the order of the number of objective functions. We present some empirical studies on this conjecture through the analysis of several well-known problems in multi-objective design. We extend the work done in [Mukerjee and Dabbeeru, 2009] to problems whose pareto-front is not a single continuous manifold, but may consist of several separate manifolds corresponding to clusters over the non-dominated space. For this purpose, we develop a variant of DBSCAN that obtains clusters on the non-dominated solutions obtained from multi-objective optimization, and illustrate the process of learning chunks as low-dimensional manifolds for each cluster. The design ramifications of these chunks are also highlighted for a electromechanical design problem.

Chapter 1

Introduction

1.1 Expert behaviour

An expert in a task domain can think effectively about problems in that domain. It is not only the general abilities such as memory and intelligence or talent that lead to expertise, rather extensive practice and exploration of the task domain are the more important factors [Ericsson, 2002]. The knowledge acquired during the exploration of the problem domain enables the expert to perceive and understand the same problem in a more effective way than a novice.

Although a novice and an expert may possess the same amount of relevant knowledge in terms of facts and formulae, the superior abilities of experts arise from the organization of the knowledge that happens as a result of experience. Experts are able to retrieve important aspects of their knowledge with little attention or effort. Their knowledge isn't just a collection of facts and propositions, instead it reflects the context of applicability, that is, knowledge is "conditionalized" on a set of circumstances [Bransford et al., 2000; Glaser, 1999]. For example an expert chess player

is able to recall meaningful board positions much more accurately than average players, while recall ability for random board positions are similar for both experts and non-experts [Chase and Simon, 1973; De Groot, 1965]. These evidences suggest that with experience and practice, people cognitively chunk the available information in their task domain into larger units based on their functional characteristics. The evidence for knowledge chunking in experts has been observed in many fields, e.g. in the game of bridge [Engle and Bukstel, 1978], in electronic circuitry [Egan and Schwartz, 1979] and computer programming [Ehrlich and Soloway, 1984].

A key aspect in chunking is abstraction. An expert is able to conceptualize the most *important* principles pertinent to solution to a problem and ignore the rest. Chunks are formed based on the functional relation of the individual elements, ignoring other less important issues. For example in chess, given a valid board position, expert players are more likely to recall only a set of “good” moves for the position, out of a very large number of possible moves.

Although design problems are different from problems in other domains [Goel and Pirolli, 1992; Schön, 1988], expert designers too rely on their experiential knowledge to guide the design process. The experience helps the designers to *frame* the problem and arrive at a principal solution early on, and identify the aspects of the design problem that need attention. Later they are able to change the early solution rather fluently and easily as difficulties are encountered.

1.1.1 A real design situation

In this work, we are considering four design problems. For one of these, the brushless DC permanent magnet motor (BDCPMM) design problem, we were able to

talk to Dr. S. P. Das who has expertise in brush-less DC motors. The problem is to design custom motors using off-the-shelf standard parts. The two objectives of the design are to minimize the production cost and to maximize the peak torque. The expression for torque produced in a BDCPMM is

$$T_p = 87300C_{tor}NR_{si}A_{wire}n_l \quad (1.1)$$

where C_{tor} is either $\frac{1}{3}$ or $\frac{2}{3}$ for Y connection and Δ connection respectively, N is the number of laminations in the stator, R_{si} is the inner radius of the stator, A_{wire} is the cross-sectional area of the wire used in the stator winding and n_l is the number of turns in the stator coil. For low cost motors, he said that without any other constraints, designs with smaller radial dimensions and thin wires for winding would be cost effective. For high end motors, however, a design with larger radial dimensions would be better in terms of cost, though the choice of wire for winding would require investigation. In both the cases similar designs would tend to vary considerably in the number of laminations in the stator. On the question of the connection type for the windings he could immediately choose the Δ connection, though the reasons for this choice were not so clear. We were surprised to note that each of these observations directly follow from our results and are discussed in section 2.7.

1.2 Modelling the chunking behaviour

The endeavour to computationally model the human learning behaviour dates back several decades. First attempts at modelling the chunking behaviour such as the

EPAM or CHREST [Feigenbaum, 1961; Gobet, 1993] tried pure and direct implementation of chunking mechanisms based on discrimination networks. An alternative approach employed in systems such as the Soar and ACT-R [Anderson, 1996] relies on a production-rule based representation of the information. A further alternative is presented by non-symbolic computational models [Elman, 2005], which propose that information should be distributed across multiple units, and no units should be dedicated to specific functions.

In the past automated design systems such as CADET [Sycara et al., 1991] and ARCHIE-II [Domeshek and Kolodner, 1991] have focused on using the knowledge stored as cases in new design problems, however they are limited in their ability to incorporate new knowledge from design experience into their knowledge base. Attempts to incorporate knowledge transfer in search oriented design systems have been made [Moss et al., 2004], however the learning is in terms of concrete design components, and carries no knowledge of the context of application. A methodology for automatically extracting innovative implicit principles of optimal design from multi-objective optimization has been proposed in [Bandaru and Deb, 2010]. In [Mukerjee and Dabbeeru, 2009] it is argued that human flexibility in symbol usage is possible because they are grounded on experience and any attempt at concrete definition will miss many of the associations the symbol has. This work is based on this view, and we propose the *chunk dimensionality conjecture* in the next section and present the results of empirical work on this conjecture in the next few chapters.

1.3 Chunk dimensionality conjecture

In many cases it has been observed that the dimensionality of the optimal solution set manifold in the decision space is the same as its dimensionality in the objective space [Mukerjee and Dabbeeru, 2009]. In this work we have tested five multi-objective optimization tasks and found that the pareto-front for these problems could be divided into clusters, most of which had manifold dimensionality comparable to the number of objectives. This evidence led us to propose the following conjecture, which we will not attempt to prove in this thesis since the constraints that the conjecture imposes are not met by most practical multi-objective optimization situations, but we will present the empirical evidence that substantiate it.

Conjecture (Chunk dimensionality conjecture). *Given a multi-objective optimization problem with a decision variable $x \in \Sigma$, where $\Sigma \subset \mathbb{R}^D$ is the decision space,*

- (a) *chunks emerge from a high-dimensional decision space Σ as clusters among the better performing combinations,*
- (b) *chunks reflect a lower dimensionality than the embedding space, i.e. chunks are manifolds of dimension d_c , $d_c < D$,*
- (c) *for multi-objective decision problems with $d + 1$ objectives ($d \ll D$), the better performing combinations are to be found on the non-dominated (pareto) frontier which is a d -dimensional manifold in the objective space, and*
- (d) *if the objective function that maps the decision space to the objective space is continuous and well-behaved, this would result in chunks that have a dimensionality $d_c = \mathbf{o}(d)$, i.e. $d_c = kd$, where $k - 1$ is vanishingly small.*

Claims a, b, and c of the conjecture are fairly well accepted, d requires further comment. An objective function f is a mapping from decision space to objective space. Let $\Phi \subset \mathbb{R}^{d+1}$ be the objective space then assuming Σ and Φ are metric spaces, an objective function f is a the mapping:

$$f : \Sigma \mapsto \Phi$$

The vector function f is continuous if and only if each component (scalar) function is continuous. A component function g of f is continuous iff:

$$\forall x, x' \in \Sigma \text{ and } \forall \epsilon \in \mathbb{R}_{\geq 0}, ;$$

$$\exists \delta \in \mathbb{R}_{\geq 0}, \text{ such that}$$

$$|x - x'| < \delta \Rightarrow |g(x) - g(x')| < \epsilon$$

i.e. two neighboring points x and x' in the decision space will map to y and y' in the objective space, which are also *near* in the objective space. Local neighborhood in the objective space may reflect either a local neighborhood in the decision space, or a combination of such neighborhoods. By well behaved we imply that the function does not oscillate very wildly or exhibit other such undesirable behaviour.

In some optimization tasks, the objective mapping function is extremely ill posed and such constraints may not hold. However in many common optimization tasks, the objective functions are quite well behaved. Table 1.1 shows that results from practical multi-objective optimization support the conjecture. In most cases at most some of the decision variables are continuous. Despite this, we can observe that for all the problems most of the clusters have a dimensionality $d_c = d$.

Design problem	D	Cont. variables	$d + 1$	Cluster dimensionality	
				Dimensionality	No. of clusters
BDCPM design	5	0	2	1	4
				2	1
Gearbox design (A)	11	10	2	1	11
Gearbox design (B)	29	10	3	2	7
Clutch brake design	5	0	2	1	5
Welded beam design	4	4	2	1	5

Table 1.1: Summary of experiments.

1.4 Dimensionality reduction

A variety of dimensionality reduction techniques are employed to find low dimensional structures [Van Der Maaten et al., 2007]. Linear dimensionality reduction techniques assume a linear relationship among the points in the data and try to find a low dimensional embedding for the data-set, while non-linear dimensionality reduction techniques assume that the points in the data-set lie along a manifold in the high dimensional space and try to find a low dimensional embedding for this manifold. For non-linear data-sets, the linear algorithms map smallest convex subspaces encapsulating the manifold, often of much higher dimension than the manifold itself. On the other hand, non-linear techniques embed the data-set in the manifold dimensions, whose relation to the original input dimensions may be difficult to unravel.

Principal component analysis (PCA) [Hotelling, 1933; Jolliffe, 2002] is by far the most popular unsupervised linear dimensionality reduction technique. PCA constructs a low dimensional representation of the data that describes as much of the variance in the data as possible. This is done by finding a linear basis of reduced dimensionality for the data using singular value decomposition techniques, in which the amount of variance in the data is maximal. The amount of variance along a

reduced dimension is its *explained variance*. We use PCA to uncover the implicit common principles that the designs in a chunk share, and to obtain a low dimensional representations of the chunks.

Nonlinear dimensionality reduction techniques fall in one of two categories: (1) techniques that attempt to preserve the local properties of the original data in the reduced dimensionality representation, and, (2) techniques that aim to preserve the global properties of the data in the low-dimensional representation. *Locally linear embedding* (LLE) [Saul and Roweis, 2003] is a major representative of local techniques. LLE models the manifold as a collection of linear patches and attempts to characterize the geometry of these linear patches. To do so it attempts to represent each point x_i as a weighted, convex combination of its k nearest neighbors. A d dimensional configuration is then found whose local geometry is characterized well the weight matrix obtained earlier. LLE has been used in [Mukerjee and Dabbeeru, 2009] to obtain a low dimensional representation for chunks of *good* designs.

Isometric feature mapping or Isomap [Tenenbaum et al., 2000] is a global technique for manifold learning that first captures the intrinsic geometry of the data in geodesic manifold distances between all pairs of data. Geodesic distances are estimated by first constructing a nearest neighbor graph in which there is an edge between every point and its k nearest neighbors or points within ϵ radius. The geodesic distance between two points is then assumed to be the shortest path distance between them in the graph so constructed. The manifold distance matrix is then used to obtain distance preserving low dimensional representation for the data using multi-dimensional scaling. The error introduced in reverse-mapping an Isomap embedding back to the original space is known as the residual error. Com-

paring the residual error for Isomap embedding of different dimensions gives a way of estimating the manifold dimensionality of the input data [Martin and Backer, 2005]. We use Isomap to estimate the dimensionality of the chunk manifolds by comparing the residual variance for embeddings of different dimension.

Chapter 2

Proposed Methodology

In this chapter we present the methodology that we are proposing for learning chunks of good designs. We give the multi-objective formulation for the the brush-less DC permanent magnet motor design problem which was discussed in section 1.1.1 earlier, and apply our procedure on this problem as we elucidate each step. The brush-less DC permanent magnet design problem was first analyzed in [Chidambaram and Agogino, 1999] as discrete optimization problem to illustrate catalog-based customization. The same problem was analyzed in [Deb and Sindhya, 2008] to discover innovative design principles from the optimization results. We borrow the same problem formulation as in [Deb and Sindhya, 2008] for our study.

2.1 Overview of the chunking process

Figure 2.1 shows the steps of our proposed chunking process. The first stage in the process is obtaining a set of optimal solutions using multi-objective optimization. The second step is the manifold estimation of the optimal solutions obtained in the first step. The next step, clustering the optimal solutions, is the most important

step in chunking procedure. In this step the optimal solutions are clustered using a density based partitioning algorithm in the combined objective-parameter space. Finally we model the clusters into manifolds which represent the chunks of similar optimal solutions.

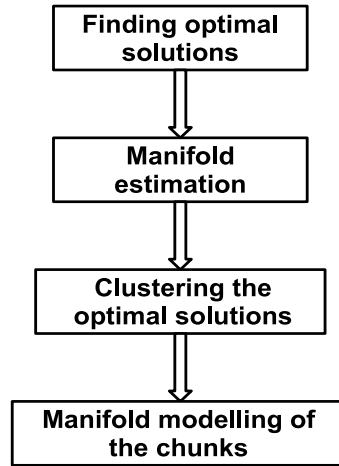


Figure 2.1: Overview of the chunking process.

2.2 Obtaining set of optimal designs

Since we want to learn chunks of “good” solutions we must have a way of characterising good solutions. In context of design, only the best possible designs may be considered as good, a designer who produces feasible but sub-optimal designs can not be called an expert. A human designer during her routine exploration of a design problem comes across some designs that are superior to all others. These are the designs that she is more likely to retain in her memory as a chunk. In effect, she is searching for optimal designs which when found are retained. The first step of our procedure to learn patterns of good designs is to obtain a representative set of

such optimal designs. In this section, we discuss the possible techniques of obtaining such representative optimal solutions set.

2.2.1 Conventional multi-objective optimization techniques

In classical multi-objective optimization techniques the objectives are combined to form one single objective using some knowledge of the problem being solved. The optimization of the single objective-results in a single pareto-optimal solution.

One of the classical methods of multi-objective optimization is the method of objective weighting [Deb, 2001; Srinivas and Deb, 1994]. Multiple objectives are combined into an overall objective by assigning fractional weights to the individual objective functions which sum to one. The optimal solution is controlled by the weight vector.

In the method of distance functions, the single objective is derived from multiple objectives using a demand level vector $\bar{\mathbf{y}}$ as follows:

$$Z = \left[\sum_{i=1}^N |f_i(\mathbf{x} - \bar{y}_i)| \right]^{1/r}, \quad 1 \leq r < \infty \quad (2.1)$$

where $\mathbf{x} \in \mathbf{X}$, the feasible region. An arbitrary demand level vector may result in a objective that doesn't have an optimal solution, hence the decision maker must have a thorough knowledge of individual optima of the objectives prior to the selection of the demand level vector.

The Min-Max formulation method is different in principle from the above two

methods in that it attempts to minimize the relative deviations from the individual optima, that is to say, it tries to minimize the objective conflict.

The most important drawback of these classical multi-objective optimization methods with respect to our objective is that they yield only one pareto-optimal solution in a single run. Secondly, all of them require some knowledge of the problem, such as the weight vector or demand vector. Moreover it is not possible to find all pareto-optimal solutions in some non-convex multi-objective optimization problems.

2.2.2 Evolutionary algorithms for multi-objective optimization

Evolutionary approaches to multi-objective optimization are much more suited to the purpose of generating a representative optimal solution set, as they can search the solution space in parallel for optimal solutions, and produce a set of pareto-optimal solutions in a single run.

Evolutionary algorithms try to mimic the process of natural selection, whereby only the specimens most adapted to the environment survive and evolve. All the algorithms start with randomly chosen population of feasible solutions and search and reproduce the best solutions in the population. This search and reproduce procedure is repeated for many generations, until the next generation doesn't result in significant improvements. The first practical evolutionary multi-objective optimization algorithm was the Vector Evaluated Genetic Algorithm (VEGA) [Schaffer, 1985]. One of the problems with VEGA is its bias towards some pareto-optimal solutions. A non-dominated sorting procedure was proposed by Goldberg to overcome this drawback, in which a ranking selection method is used to emphasize good points

and a niche method is used to maintain stable sub-populations of good points. One algorithm that implements this non-dominated sorting procedure is the NSGA [Srinivas and Deb, 1994].

This algorithm was further improved upon, and a new algorithm NSGA-II [Deb et al., 2002] was proposed which had better and faster convergence towards the true pareto-front due to a fast non-dominated sorting approach. Since it was first proposed, the NSGA-II has found application in a wide variety of engineering optimizations and design problems. Several studies have been published illustrating the versatility of NSGA-II in a wide variety of situations. All these advantages that NSGA-II has over other algorithms, and its proved versatility make it an ideal candidate for the purpose of generating a set of pareto-optimal solutions

2.2.3 The BDCPM motor design problem

We now introduce the multi-objective optimization formulation for the BDCPM (brush-less DC permanent magnet motor) design problem and derive a set of pareto-optimal solutions using NSGA-II. A BDCPM motor comprises of an outer stator assembly with windings on a frame and inner rotor assembly having permanently mounted magnets, shown in Figures 2.2 and 2.3. Several variants of this basic design exist. The original study [Chidambaram and Agogino, 1999] considered 24 slots 4 pole machine.

The multi-objective optimization problem

Figure 2.3 shows the rotor assembly of a BDCPM motor. A ring-magnet is bonded onto a stepped shaft that fits in the bore of the stator assembly. The total length of

the motor is L_{sh} . The shaft length is equal to the sum of stack length and the side clearances, $L_{sh} = L + 2L_{cl}$. The side clearance values L_{cl} , and other parameters for different lamination types are listed in Table 2.1.

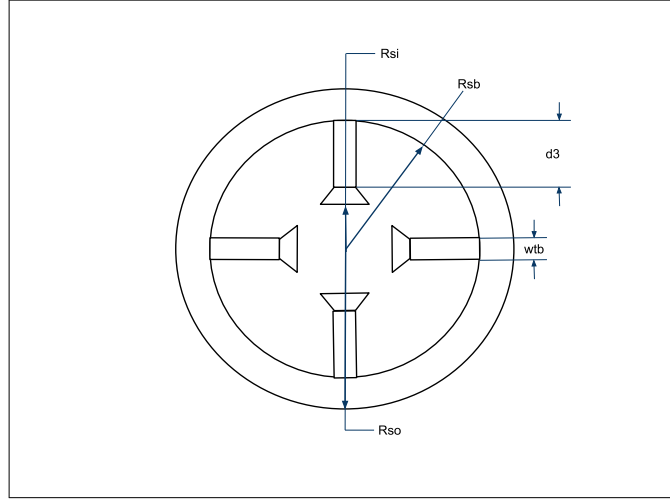


Figure 2.2: BDCPMM Stator assembly.

L_{type}	R_{si} (mm.)	d_3 (mm.)	w_{tb} (mm.)	w_{bi} (mm.)	L_{cl} (mm.)	R_{sh} (mm.)	R_{st} (mm)
X	21.90	12.0	2.39	5.23	4.215	4.11	0.016
Y	22.22	15.1	2.39	5.23	4.265	4.11	0.016
Z	25.40	15.1	2.80	5.50	4.775	4.45	0.019

Table 2.1: Dimensions of the BDCPM motor family

Five design variables are considered for the design process, other design parameters are fixed to values that are convenient for the manufacturer. The design variables are:

1. Number of laminations $n_l \in [44, 45, \dots, 132]$
2. Number of turns in each coil $N \in [20, 21, \dots, 80]$
3. $L_{type} \in [X, Y, Z]$ is one of the three types of laminations allowed to build the motor
4. $M_{ph} \in [Y, \Delta]$ is one of the two types of electric connections and

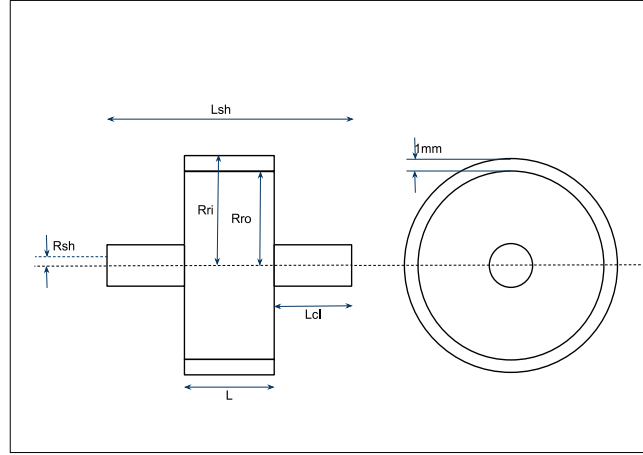


Figure 2.3: BDCPMM Rotor Assembly.

5. $A_{gauge} \in [16, 16.5, 17, \dots, 23.5]$ is one of the 16 wire gauges used in the windings.

All the design variables are discrete in nature, of which L_{type} and M_{ph} take 3 and 2 values respectively. n_l and N are integer valued variables while A_{gauge} takes 16 equally spaced values in its domain.

The cost expression for the first objective includes the material and estimated production cost. Each term in the expression is obtained from the regression analyses of data obtained from practice. The detailed procedure describing the derivation of the cost terms and torque expressions can be found in [Chidambaram and Agogino, 1999].

The actual multi-objective optimization problem is formulated as follows:

Minimize:

$$C_{total}(n_l, N, L_{type}, M_{ph}, A_{wire}) = \begin{cases} 2.6\left(\frac{n_l}{44}\right)^{0.25}, & \text{for } L_{type} = \mathbf{X} \\ 0.38 + 2.42\left(\frac{n_l}{44}\right)^{0.37}, & \text{for } L_{type} = \mathbf{Y} \\ 1.07 + 1.83\left(\frac{n_l}{44}\right)^{0.58}, & \text{for } L_{type} = \mathbf{Z} \end{cases}$$

$$+ [\{0.026, 0.028, 0.03\} n_l \cong L_{type} = \{\mathbf{X}, \mathbf{Y}, \mathbf{Z}\}]$$

$$+ \frac{n_l}{44} + 9.8 \times 10^5 A_{wire} N [5.8 \times 10^{-4} n_l + \frac{\pi}{2} \{w_{tb} + \frac{5\pi}{12} (R_{si} + \frac{d_3}{2})\}]$$

$$+ 0.3035 + 0.876 N [5.8 \times 10^{-4} n_l + \frac{\pi}{2} \{w_{tb} + \frac{5\pi}{12} (R_{si} + \frac{d_3}{2})\}]$$

$$+ \pi [\{31.2(R_{si} + d_3 + w_{bi}) + 0.0312\} \{L_{sh} - 5.08 \times 10^{-3}\}]$$

$$\begin{aligned}
& + \left\{ \begin{array}{ll} 0.7 \left(\frac{5.8 \times 10^{-4} n_l + 0.0792}{0.1047} \right)^{1.21}, & \text{for } L_{type} = \mathbf{X} \\ 0.54 + 0.26 \left(\frac{5.8 \times 10^{-4} n_l + 0.0802}{0.1057} \right)^{3.74}, & \text{for } L_{type} = \mathbf{Y} \\ 0.58 + 0.32 \left(\frac{5.8 \times 10^{-4} n_l + 0.0905}{0.1160} \right)^{4.23}, & \text{for } L_{type} = \mathbf{Z} \end{array} \right. \\
& + 9 + \left\{ \begin{array}{ll} 1.3 \left(\frac{5.8 \times 10^{-4} n_l + 0.0792}{0.1047} \right)^{0.38} \left(\frac{n_l}{44} \right)^{0.18}, & \text{for } L_{type} = \mathbf{X} \\ 1.6 \left(\frac{5.8 \times 10^{-4} n_l + 0.0802}{0.1057} \right)^{0.38} \left(\frac{n_l}{44} \right)^{0.40}, & \text{for } L_{type} = \mathbf{Y} \\ 1.8 \left(\frac{5.8 \times 10^{-4} n_l + 0.0905}{0.1160} \right)^{0.41} \left(\frac{n_l}{44} \right)^{0.52}, & \text{for } L_{type} = \mathbf{Z} \end{array} \right. \\
& + (1.536 R_{st} - 6.26 \times 10^{-3}) n_l + (2.014 R_{si} - 8.21 \times 10^{-3}) n_l \\
& + \pi \{ 29.4 (R_{si} - 7.25 \times 10^{-3})^2 n_l + 1.014 \times 10^5 R_{sh}^2 L_{cl} \} \\
& + 0.1862 + 5.31 \times 10^4 \pi [R_{st}^2 L_{st} - 2 R_{sh}^2 L_{cl} - 5.8 \times 10^{-4} n_l (R_{si} - 7.25 \times 10^{-3})^2] \\
& + 0.01 + 0.1473 \pi (R_{si} - 7.25 \times 10^{-3}) n_l + \left\{ \begin{array}{ll} 1.2, & \text{for } T_p \leq 3.5 \\ 1.6, & \text{for } T_p > 3.5 \end{array} \right. \\
& + [\{0.5, 0.55, 0.60\} \cong L_{type} = \{\mathbf{X}, \mathbf{Y}, \mathbf{Z}\}] \\
& + \left\{ \begin{array}{ll} 3.2 \left(\frac{5.8 \times 10^{-4} n_l + 0.0792}{0.1047} \right)^{-1.92} \left(\frac{n_l}{44} \right)^{0.92}, & \text{for } L_{type} = \mathbf{X} \\ 3.4 \left(\frac{5.8 \times 10^{-4} n_l + 0.0802}{0.1057} \right)^{-1.92} \left(\frac{n_l}{44} \right)^{1.06}, & \text{for } L_{type} = \mathbf{Y} \\ 3.7 \left(\frac{5.8 \times 10^{-4} n_l + 0.0905}{0.1160} \right)^{-3.25} \left(\frac{n_l}{44} \right)^{1.60}, & \text{for } L_{type} = \mathbf{Z} \end{array} \right.
\end{aligned}$$

Maximize:

$$T_p = 87300 C_{tor} N R_{si} A_{wire} n_l$$

where

$$C_{tor} = \left\{ \frac{2}{3}, \frac{1}{3} \right\} \cong M_{ph} = \{\mathbf{Y}, \Delta\}$$

Subject to

$$g_1 : T_p \geq 0.83$$

$$g_2 : T_p \leq 5.27$$

$$g_3 : A_{wire} N \leq \{150, 240, 280\} \times 10^{-7} \cong L_{type} = \{\mathbf{X}, \mathbf{Y}, \mathbf{Z}\}$$

Constraints g_1 and g_2 bound the peak torque in the desired range. Constraint g_3 represents the winding constraints that prevent magnetic saturation and demagnetization.

NSGA-II formulation and the pareto-front

Since all the variables are discrete, we use binary representation for them in NSGA-II. We modify the population initialization routine and the EA operators (recombination and mutation) to maintain the discrete nature of the variables. For the two integer valued variables, n_l and N , discrete version of the Simulated Binary Crossover (SBX) and Polynomial Mutation [Deb, 2001] are used. Since M_{ph} and

L_{type} take two and three values only, they are only subjected to mutation. Using this representation NSGA-II yields 199 non-dominated solutions, plotted in figure 2.4 in red circles.

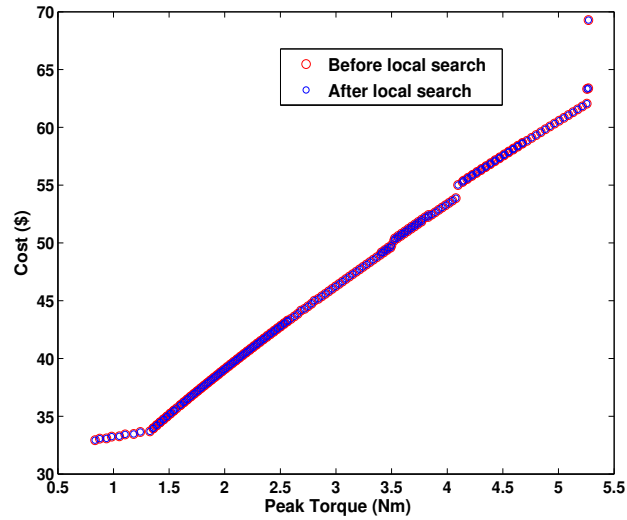


Figure 2.4: Pareto-front for the BDCPMM design problem. The red circles represent the design results of NSGA-II and the blue circles represent the design results of the local search. Pareto-front obtained after the local search overlaps with the NSGA-II result.

2.2.4 Local search procedure

Although an evolutionary procedure may find a set of near optimal solutions quickly, the convergence to the true pareto-front might take indefinitely long. To improve convergence characteristics of the evolutionary multi-objective algorithms they are often complemented with a local search procedure. Local search is more important in case of multi-objective optimization problems which involve discrete variables. We have used a generic local search procedure which evaluates random solutions in the vicinity of the near-optimal solution under consideration. The input for local search are the set of near optimal solutions S , number of local search iterations

to perform n , number of local solutions to evaluate for each near optimal solution h , and the search width factor σ . Algorithm 1 gives the pseudo-code for our local search procedure.

```

input :  $(S, n, h, \sigma)$ 
output: Set of optimal solutions  $P$ 
 $P = \phi$ ;
for  $i = 0$  to  $n$  do
  for each  $s \in S$  do
    add  $s$  to  $P$ ;
    find  $k$  nearest neighbors of  $s$ ;
    find the range of values each variable takes in the  $k$ -neighborhood;
    for  $j = 0$  to  $h$  do
      generate a candidate random solution  $p$  having variable values in
      the neighborhood range scaled by  $\sigma$ ;
      add  $p$  to  $P$ ;
    end
  end
  perform non-dominated sorting on  $P$  and discard dominated solutions;
   $S = P$ ;
end
return  $P$ ;

```

Algorithm 1: Local search procedure.

For the BDCPMM design problem, however, local search doesn't result in much improvement since we ran NSGA-II for a large number of generations(3000) which would have been sufficient for convergence to the actual pareto-front. Performing local search on the NSGA-II output for BDCPMM design problem increases the number of optimal solutions to 201 from 199.

2.3 Estimating the pareto-front manifold

We apply Isomap and PCA to estimate the dimensionality of the pareto-front manifold and learn global properties of the set of the good designs. After clustering this process will again be applied to the obtained chunks to elicit the chunk local

properties and capture their semantics.

2.3.1 Isomap Residual Variance for the Pareto-front

On performing Isomap on the pareto-front data of BDCPMM design we obtain a residual variance plot shown in figure 2.5a. Since the residual variance shows a sharp decline going from one dimensional embedding to two-dimensional embedding, we can expect the pareto-front to be composed of sub-manifolds of one or two dimensions. There is an unexpected increase in the residual variance for three dimensional embedding, though it drops down to the previous level for four dimensional embedding.

2.3.2 PCA explained variance

The PCA analysis of BDCPMM pareto-front gives four principal components. The explained variance for principal components of the BDCPMM pareto-front data are plotted in figure 2.5b. The first principal component has explained variance of more than 0.95, while the second principal component has explained variance of 0.021. The third and fourth principal components have negligible explained variances. Table 2.2 shows the first two principal component vectors which reveal that the pareto front is a manifold embedded in the plane of first two design dimensions, namely number of laminations and number of turns in the stator windings.

	n_l	N	L_{type}	M_{ph}	a_{gauge}
First PC	0.9993	0.0370	0.0072	0	-0.0024
Second PC	-0.0367	0.9943	0.0165	0	0.0985

Table 2.2: First two principal components of the BDCPM data. Number of laminations is the dominant variable in the first principal component.

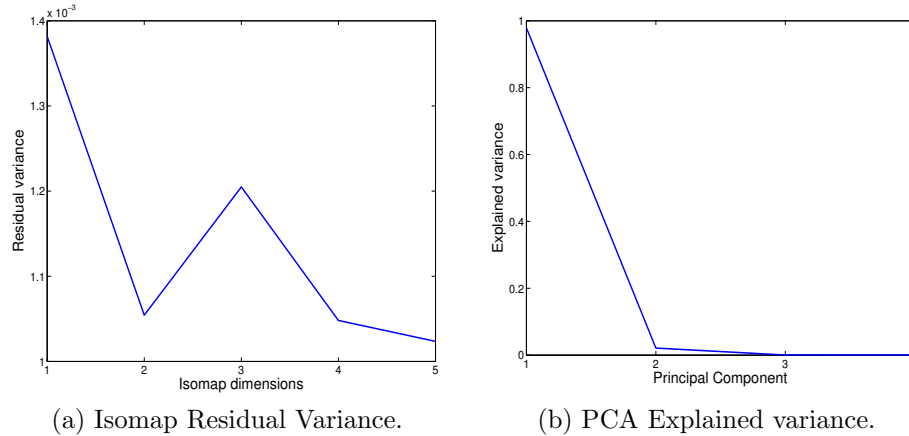


Figure 2.5: Isomap and PCA results. The residual variance plot shows the largest drop for the second dimension indicating a manifold of dimension one. First principal component has an 95% explained variance indicating a linear dimensionality of one.

2.4 Clustering the pareto-front

Our chunking process aims to identify solutions with similar characteristics and group them together. Those solutions of a multi-objective optimization which are close together in the combined objective-decision variables space, can be expected to have some shared characteristics. Clustering the pareto-front is a way to group similar solutions together.

Figure 2.6 shows the 3-dimensional Isomap embedding of the BDCPMM pareto-front. The embedding algorithm used in the Isomap is the classical multi-dimensional scaling (cMDS), which aims to preserve the inter-point distances in the reduced dimension. The 3-dimensional Isomap embedding is thus an approximate modelling of the data set in the original parameter space, with distances along the manifold preserved. This figure clearly suggests that the BDCPMM pareto-front is composed of one-dimensional manifolds, some of which run parallel to each other. Our aim in clustering the pareto-front is to recover these lower dimensional sub-manifolds.

We use an algorithm that is based on concepts of density-based partitioning

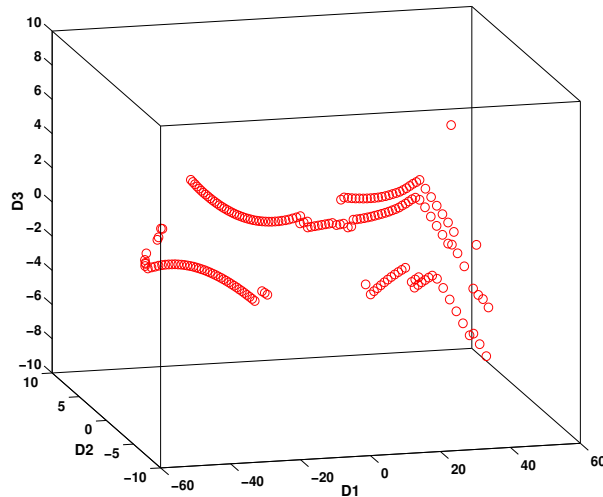


Figure 2.6: 3-dimensional Isomap Embedding of the BDCPMM problem Pareto-front, showing several one-dimensional manifolds.

to cluster the Pareto-fronts. One advantage of density based clustering is that it is able to discover clusters of arbitrary shapes. DBSCAN is a major representative of density based clustering. DBSCAN works by identifying *core objects* in the data, those points in the data-set that have at least $MinPts$ points within their ϵ -neighborhoods. Two points x and y have *density connectivity* between them if there exists a sequence of core objects between x and y such that each next belongs to an ϵ -neighborhood of its predecessor. Two points with density connectivity between them belong to the same cluster. $MinPts$ and ϵ are the only two parameters that this algorithm requires.

Our variant utilizes the concept of density connectivity, but we do not attempt to explicitly identify the core objects. Instead, density connectivity between points is determined by building a graph based on density. Initially, all the points in the data-set are nodes in the graph. We start by finding n_{max} nearest neighbors of each point in the data-set and sort these lists of nearest neighbors in the increasing order

of their distances from the point. At the beginning of each iteration, the average edge weight (ad_c) and edge weight standard deviation (σ_c) are calculated for each component. We connect each point p with the first point in its nearest neighbor list only if its distance from the point is within $k\sigma_c + ad_c$ for the component c to which p belongs, or else if average edge weight in the component is zero (p is not connected to any other node), then average global inter-point distance is used in place of ad_c . If an edge is added then the nearest neighbor is popped from the list, otherwise we don't process p in subsequent iterations, as it means that either p is an outlier or a boundary point of the cluster. This process is repeated for n_{max} iteration, by which time all the nearest neighbors lists will be empty. The pseudo-code of the algorithm is shown in the listing 2.

Our variant differs from DBSCAN in two important aspects. First, our algorithm allows for clusters with different intra-cluster point densities, since the decision that a point belongs to a cluster is based on its distance from the nearest point in the cluster, and the average edge weight within the component. DBSCAN may fail to identify a cluster with lower point density because of a low ϵ parameter. Secondly in place of the ϵ parameter we have n_{max} parameter giving the number of nearest neighbors that must be found for each point. The number and size of clusters depends on the parameter k .

The clustering of the BDCPMM pareto-front results in five clusters which are shown in the objective space plot in figure 2.7. Figure 2.8 shows the clusters with respect to the number of laminations (n_l) and the number of turns (N) in the stator windings. Other design variables are indicated alongside the clusters.

```

input : ( $D, n_{max}, k$ )
output: A cluster labelling for each data-point

find  $n_{max}$  nearest neighbors of each point arranged in non-decreasing order of
the distance;
 $ad_g$  = average inter-point distance in the data set;
 $\sigma_g$  = standard deviation of distances in the data set;
 $G = \{V, E\}$  be a graph where  $V$  is the set of all data-points and  $E = \phi$ ;
for  $i = 1$  to  $n_{max}$  do
    for each connected component  $c$  in  $G$  do
        | find out the average and standard deviation of edge weights in  $c$ ;
    end
    for each point  $p$  in the data set do
        | if  $p$ 's nearest neighbor list is empty then
        | | continue;
        | end
        |  $c$  = component label of the component  $p$  belongs to;
        |  $q$  = the first point in the nearest neighbor list of  $p$ ;
        |  $d$  = distance to first point in ordered nearest neighbors list of  $p$ ;
        | if  $i = 0$  then
        | |  $thresholdDistance = ad_g + k\sigma_g$  ;
        | else
        | |  $thresholdDistance = ad_c + k\sigma_c$  ;
        | | /*  $ad_c$  and  $\sigma_c$  are the average and standard deviations of
        | | edge weights in  $c$  */
        | end
        | if  $d > thresholdDistance$  then
        | | /* if  $i = 0$  then  $p$  is either an outlier or boundary */
        | | remove all points from the ordered nearest neighbors list of  $p$ ;
        | else
        | | if  $(q, p) \notin E$  then
        | | | add  $(p, q)$  to  $G$ ;
        | | | remove  $q$  from ordered nearest neighbors list of  $p$ ;
        | | end
        | end
    end
end
end
label points in  $V$  with the component they belong to;
return  $V$ ;

```

Algorithm 2: Clustering algorithm to obtain low-dimensional clusters

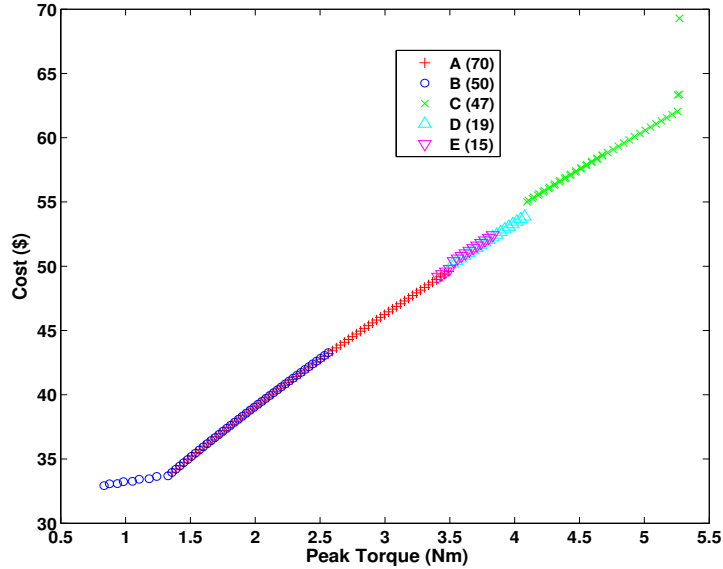


Figure 2.7: Clusters in the BDCPMM pareto-front. Clusters **A** and **B** are composed of low power and cost designs while Cluster **D** as the high-power and cost designs.

2.5 Manifold modelling of the clusters

In this step we apply manifold learning and linear dimensionality reduction to the clusters obtained in the previous step in an attempt to uncover the nature of the clusters. We again use Isomap to uncover the shape and dimensionality of each cluster and PCA to elicit shared design characteristics of solutions in each cluster. However, since the clusters may be linear patches of the whole manifold, PCA analysis reveals more useful information than Isomap analysis. All the clusters except **C** have negligible or zero residual variance for one dimensional Isomap embedding, indicating that they are one-dimensional. Residual variances for **C** are plotted in figure 2.9. The residual variance shows a large drop from two dimensional Isomap embedding to three dimensional Isomap embedding, after which it is flat. This means that cluster **C** has intrinsic dimensionality of two. Looking at figure 2.7 and 2.7 reveals why it is so. The cluster **C** has higher torque motors, but unlike other

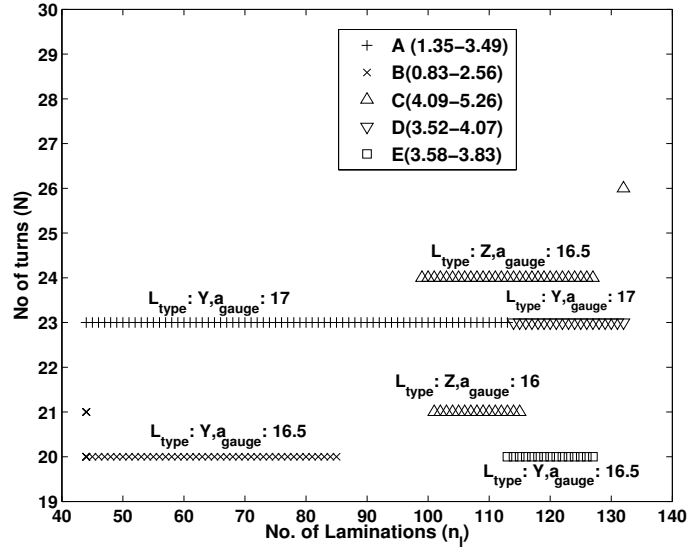


Figure 2.8: Clusters in the n_l - N subspace. All the clusters appear as lines in the parameter space except the cluster of high-end designs **D** which appears as two separate lines.

clusters which have the same number of turns in the stator windings, it has two types of designs with 21 and 24 turns. The wire gauges are also distinct in these two types of designs, those with 21 turns use 16 gauge wire and those with 24 turns use 16.5 gauge wire.

These findings are in accordance with the *chunk dimensionality conjecture* proposed in section 1.3, whereby chunks are manifolds in the decision space of dimensionality less than the dimensionality of the decision space. The dimensionality of these manifolds in the decision space is also less than or equal to the dimensionality of the objective space, which is two in this case. Four clusters have manifold dimensionality of less than two and only one has two dimensions.

Table 2.3 shows the explained variances for principal components of each of the clusters. The explained variances corroborate the observations made in the Isomap analysis that four of the clusters are one dimensional and only one cluster (**C**) is

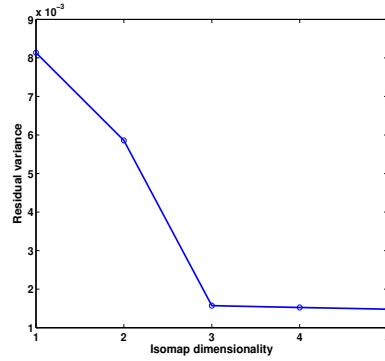


Figure 2.9: Residual variance for the cluster **C**. The largest drop in the residual variance is for the 3-dimensional Isomap embedding, indicating a two dimensional manifold.

two dimensional. Table 2.4 lists the first principal components of the clusters **A**, **B**,

D **E**. Since cluster **C** is two dimensional, its two principal components are shown.

A	100	0	0	0	0
B	99.857	0.119	0.02	0	0
C	70.82	29.17	0.01	0	0
D	100	0	0	0	0
E	100	0	0	0	0

Table 2.3: Explained variances for the principal components of the clusters. For clusters **A**, **B**, **D** and **E** there is only one significant principal component. Cluster **D** has two significant principal components.

	n_l	N	L_{type}	M_{ph}	a_{gauge}
A	1	0	0	0	0
B	0.999	0.0077	0	0	-0.019
C	0.789	0.610	0	0	0.066
	-0.613	0.785	0	0	0.074
D	1	0	0	0	0
E	1	0	0	0	0

Table 2.4: Principal components for the clusters. For **A** **B** **D** and **E** the principal component is along the direction of n_l . Cluster **C** has two components which are in the $n_l - N$ plane.

2.6 Design implications

There are a number of important facts that can be observed from the PCA analysis of the clusters. Firstly, as is evident in the tables 2.2 and 2.4, number of laminations is the most important design parameter, as this is the variable that has the largest weights in the first principal component of the whole pareto-front and the clusters. This parameter can be changed for inducing a small amount of change in the peak torque and cost.

The connection type (M_{ph}) has no contribution whatsoever in the principal components of the pareto-front and the clusters. This means that all the optimal designs must be using the same connection type. Only *Y* connection type is used in all the designs.

Although the variable L_{type} has no weight in the principal components of the clusters, it has some contribution to the principal component of the overall pareto-front. This means that lamination type is the same for all the designs in a single cluster, though it may vary across the pareto-front. Only two out of three lamination types are used in all the designs. Lamination type *Y* is used in all the clusters except **C**, in which lamination type *Z* is used. Looking at figure 2.8 we can see that cluster **C** occupies the highest torque and cost range of the pareto-front. This implies that lamination type *Z* is only useful for high end designs of BDCPM motors.

Number of turns (N) has a small weights in the principal components of the pareto-front. The principal components of clusters **A**, **D** and **E** have zero weights for N , and **B** and **C** have non-zero weights. Incidentally, these are also the only clusters that have non-zero weights for the variable a_{gauge} . This suggests that these two variables are not independent of each other in the optimal designs. The number

of turns tend to stick to the lower region (20–24) of the range of possible values (20–80). Also higher cross-section wires (lower a_{gauge}) among the available choices are used in the optimal designs

2.7 How the chunks relate to experts' intuition

The observations made in the previous section are in confirmation of the expert's intuition of the design of a BDCPM motor, discussed in the section 1.1.1. To the expert it was 'natural' that if higher peak torque is desired then number of laminations have to be increased proportionately. Increasing the number of turns in the stator winding also increases the torque, but in much larger quanta's. Also, since non-customizable lamination designs are used, it may not be possible to increase the number of turns, the reason being that the size of the groove (d_3 in figure 2.2 and table 2.1) that accommodates the windings is fixed for each lamination type. So to increase the number of turns either a lamination type with larger radial dimensions must be used or a wire with smaller cross-section must be chosen for the windings.

All these trade-offs can be clearly observed in play in the cluster of high-end designs (C) where the lamination type with largest radial dimension is used and two combinations of number of turns and wire cross-section are employed. One set of designs has 21 turns with 16 gauge wire and the other has 24 turns with 16.5 gauge wire.

When the expert had to choose the connection type in the winding, he chose the Y type connection, although reasons for the choice were not immediately clear. After some back of the envelop calculations, the explanation he gave was that since the wire cross-section is fixed, the maximum current that can flow in the stator

winding is limited and Y connection type allows for a higher power factor. Indeed, all the optimal designs in our results have a Y connection type.

Chapter 3

The Gearbox Design Problem

The Gear Train design problem is a classic multi-objective optimization problem, for it encompasses a number of the challenges that a practical real world optimization task may present. It involves a number of variables of different types and a large number of equality and inequality constraints. A number of objectives can be considered for the optimization. This problem was analyzed in [Jain and Agogino, 1990] to explain the design process from an optimization perspective, and in [Deb and Jain, 2003] to prove the applicability of NSGA-II in complex optimization tasks. We analyze the pareto-optimal solutions of this design-optimization problem to discover chunks of optimal designs using our proposed procedure.

3.1 Problem description

The objective of multi-speed gearbox design is to obtain different specified speeds in the output shaft with fixed input shaft angular velocity. The two-dimensional layout of a 18-speed gearbox with 18 gears is shown in figure 3.1. Gears in shafts 1, 3 and 5 can be translated to mesh with corresponding gears in shafts 2, 4 and 6

to obtain desired output speed. The input shaft speed is kept fixed at 1400 rpm. With a ratio of 1.14 between consecutive output speeds, the i^{th} desired output speed would be $1400/(1.14)^{i-1}$. The lowest desired speed is thus $1400/(1.14)^{17} = 150$ rpm.

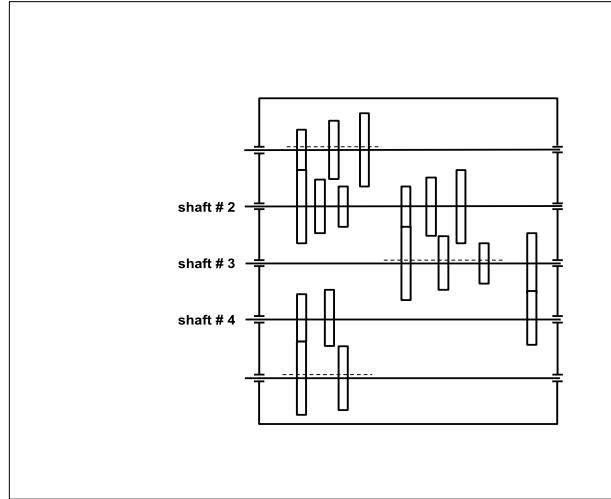


Figure 3.1: 18-speed Gearbox layout. Gears in shafts 1, 3 and 5 can be translated to mesh with corresponding gears in shafts 2, 4 and 6 respectively. With three different possible meshing of gears between shafts 1-2, 2-3 and two in shafts 2-4, 18 different output speeds are possible.

A number of objectives can be considered for the design of the gear train design optimization, e.g.:

- (i) minimization of the gear material used,
- (ii) maximization of the power delivered,
- (iii) minimization of the error between desired and achieved output speeds.
- (iv) minimization of the centre distance between input and output shaft,

We will consider only the first three objectives in this study.

There are 29 variables over which an optimization problem can be defined:

- (i) thickness of gear-pairs, t_i for $i = 1, 2, \dots, 9$

- (ii) teeth module, the same for all the gear pairs,
- (iii) the power delivered p ,
- (iv) the number of teeth in pinion of the i^{th} gear-pair,
- (v) the number of teeth in wheel of the i^{th} gear-pair, for $i = 1, 2, \dots, 9$,

To have varying number of teeth in the gears, the following constraints must be satisfied by all feasible gearbox designs:

1. All mating gear-pairs on two shafts should have the same centre distances as the shaft distances.
2. Since we have a two-dimensional gear-train layout, no gear should interfere with any shaft.
3. Maximum gear ratio in any gear pair should not exceed a limit r^{max} .

The multi-objective formulation for the full problem is as follows:

$$\text{Maximize } f_1 = p, \quad (3.1)$$

$$\text{Minimize } f_2 = \frac{\pi m^2}{4} \sum_{i=1}^G (n_{p_i}^2 + n_{w_i}^2) t_i \quad (3.2)$$

$$\text{Minimize } f_3 = \max_{1 \leq k \leq 18} |\Omega_k - \Omega_k^I| \quad (3.3)$$

Subject to,

$$\sigma_{b_i} \leq S_b \quad \text{for } i = 1, 2, \dots, G, \quad (3.4)$$

$$\sigma_{w_i} \leq S_w \quad \text{for } i = 1, 2, \dots, G, \quad (3.5)$$

$$t^{(L)} \leq t_i \leq t^{(U)} \quad \text{for } i = 1, 2, \dots, G, \quad (3.6)$$

$$p^{(L)} \leq p \leq p^{(U)} \quad (3.7)$$

$$n_1 + n_2 = n_3 + n_4 = n_5 + n_6, \quad (\text{between shaft 1 and 2})$$

$$n_7 + n_8 = n_9 + n_{10} = n_{11} + n_{12}, \quad (\text{between shaft 2 and 3})$$

$$n_{15} + n_{16} = n_{17} + n_{18} \quad (\text{between shaft 4 and 5}) \quad (3.8)$$

$$\begin{aligned} n_2 &\leq n_7 + n_8, & n_4 &\leq n_7 + n_8, \\ n_6 &\leq n_7 + n_8, & n_7 &\leq n_1 + n_2, \end{aligned} \quad (3.9)$$

$$\begin{aligned} n_9 &\leq n_1 + n_2, & n_{11} &\leq n_1 + n_2, \\ n_8 &\leq n_{13} + n_{14}, & n_{10} &\leq n_{13} + n_{14}, \end{aligned} \quad (3.10)$$

$$\begin{aligned} n_{12} &\leq n_{13} + n_{14}, & n_{13} &\leq n_7 + n_8, \\ n_{14} &\leq n_{15} + n_{16}, & n_{15} &\leq n_{13} + n_{14}, \end{aligned} \quad (3.11)$$

$$n_{17} \leq n_{13} + n_{14} \quad (3.12)$$

$$\begin{aligned} \frac{n_{w_i}}{n_{p_i}} &\leq r^{max} && \text{(for } i = 1, 2, \dots, G) \\ n_{w_i} &\geq n^{(L)} && \text{(for } i = 1, 2, \dots, G) \\ n_{p_i} &\geq n^{(L)} && \text{(for } i = 1, 2, \dots, G) \\ n_i &\in \mathbb{Z} && \text{(for } i = 1, 2, \dots, 2G) \end{aligned} \quad (3.13)$$

Where,

- m is the teeth module for all the gear pairs,
- n_{p_i} and n_{w_i} are the number of teeth in the pinion and wheel in the i^{th} gear-pair,
- t_i is the thickness of the gear-pair in cm.,
- Ω_k is the actual k^{th} output speed and Ω_k^I is the ideal k^{th} output speed,
- σ_{b_i} is the bending stress developed in a gear-pair calculated as follows:

$$\sigma_{b_i} = \frac{97500pk_c k_d (r_i + 1)}{a_i \omega_i t_i m r_i y_i \cos \beta} \quad (3.14)$$

- S_b ($= 2,500$ kgf/cm²) is the permissible bending stress,
- σ_{w_i} is the wearing stress developed in a gear-pair calculated as follows:

$$\sigma_{w_i} = \frac{0.59(r_i + 1)}{r_i a_i} \sqrt{\frac{97500pk_c k_d E (r_i + 1)}{\omega_i t_i \sin 2\beta}} \quad (3.15)$$

- S_w ($= 17,500$ kgf/cm²) is the permissible wearing stress,
- k_c ($= 1.5$) is the stress concentration factor,
- k_d ($= 1.1$) is the dynamic load factor,
- r_i is the transmission ratio, defined as the ratio of number of teeth in the wheel (n_{w_i}) to the number of teeth in the pinion (n_{p_i}) for the i^{th} gear-pair,

- ω_i is the angular velocity of the wheel in rpm,
- a_i is the centre distance for the corresponding gear pair given as $a_i = \frac{m(n_{w_i} + n_{p_i})}{2}$,
- y_i is the form load factor defined as $y_i = 0.52(1 + \frac{20}{n_{w_i}})$,
- β (= 20 degrees) is the pressure angle,
- E ($= 2.1 \times 10^6$ kgf/cm²) is the Young's modulus of the gear material. .

Here we will consider two instances of the problem:

- (A) 11 variable problem in which thickness of all the gear-pairs and gear-module thickness are variable and only first two objectives are considered, and
- (B) the full problem.

For the instance (A), the number of gear teeth layout is fixed to the configuration given in table 3.1. Constraints 3.8 onwards are not applicable for this instance of the problem.

i	1	2	3	4	5	6	7	8	9
n_{p_i}	20	33	28	32	34	33	33	20	26
n_{w_i}	56	43	48	37	35	36	34	54	48

Table 3.1: No. of teeth for each gear-pair in fixed layout gear-train design.

3.1.1 NSGA-II formulation and the pareto-front

We use the same NSGA-II formulation as in [Deb and Jain, 2003] to obtain the pareto-optimal solutions. For the 11 variables case we fix the gear-box layout by fixing the number of teeth in each gear to the values shown in table 3.1.

In second case we utilize the equality constraints given in equations 3.4 to eliminate five gear teeth variables. We code the gear teeth variables in such a way so

as to avoid evaluating permutations of same combinations of gears in each transmission stage. Also, we condense similar constraints into one constraint wherever possible. The probabilities of recombination and mutation operator are 0.9 and 0.3 respectively for both the optimization cases. For the 11 variables optimization, we run NSGA-II for 600 generations with a population size of 3000, and for 10000 generations with 3000 initial population for the 29 variable case. NSGA-II yields 927 points in the pareto-front for the first case. The pareto-front for the 29 variables optimization has 989 points. The pareto-fronts are shown in figures 3.3 and 3.6.

3.2 Fixed transmission ratio problem

Figure 3.9a shows the Isomap residual variances for the pareto-front of the fixed lay-out gearbox pareto-front. The largest drop in the Isomap residual variance from 3 to 4 dimensional Isomap embedding. This suggests the pareto-front may have an intrinsic dimensionality of four, and we can expect to have clusters of dimensions lower than 4.

The PCA explained variance is shown in 3.9b, which clearly shows that the pareto-front is one dimensional in the parameter space. Table 3.2 shows the weights of five variables which have the highest absolute component in in the first two principal components. The thicknesses of the gear pairs in the final transmission stage are the variables with the highest weight in the first principal component. Next are the thickness of the second and fifth gear-pairs. In the second principal component, the gear module is the most important variable. Weights of all the other variables are negligible in this principal component.

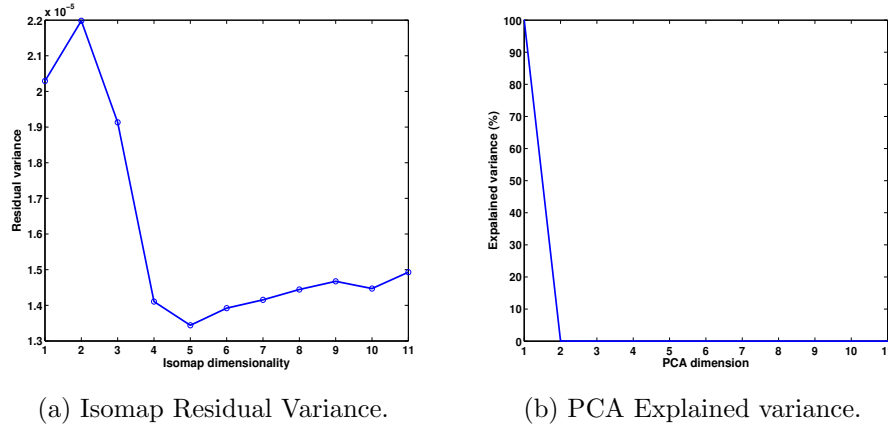


Figure 3.2: Isomap and PCA results for the fixed lay-out problem. The largest drop in residual variance is for the four dimensional Isomap embedding. The explained variance plot shows only one significant component.

First PC	(t_9) -0.8444	(t_8) -0.4617	(t_7) 0.1837	(t_6) -0.1511	(p) 0.0733
Second PC	(m) 0.9997	(t_5) -0.0124	(t_6) 0.0097	(t_9) 0.0087	(t_8) 0.0087

Table 3.2: First two principal components of the fixed layout gearbox pareto-front. t_9 and t_8 are the variables with the highest weights in the first component.

3.2.1 Clustering the pareto-front

Although it is possible to cluster the pareto-front into smaller clusters by varying the k parameter in our algorithm, we show and analyze significantly sized and smaller number of clusters for the sake of simplicity. Figure 3.3 shows the 11 clusters obtained by setting $k = 6.0$.

Isomap and PCA analysis

The Isomap residual variance for most of the clusters is negligible except for the cluster **D** for the residual variance is plotted in figure 3.4a. The largest drop in residual variance is observed for the second Isomap dimension after which it is constant, suggesting a manifold dimensionality of one. All the clusters thus have a

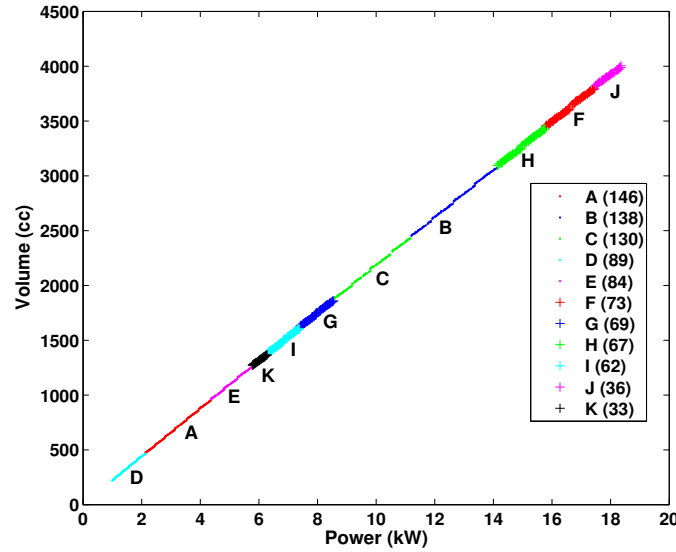
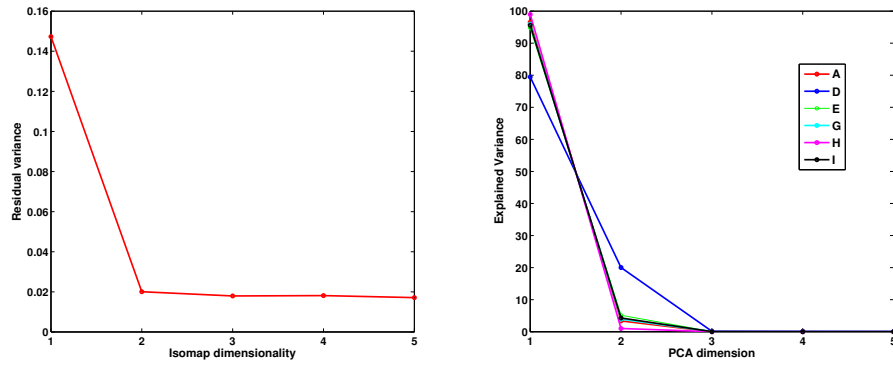


Figure 3.3: Fixed layout gearbox design pareto-front and clusters in the objective space. There are 927 solutions in the pareto-front. The clusters in the extremes of the pareto-front are small.

manifold dimensionality of one, conforming to the *chunk dimensionality conjecture*. PCA analysis shows that clusters **B**, **C**, **F**, **J** and **K** are clearly one dimensional with the first principal component of these clusters having explained variance of more than 99%. For the cluster **D**, the first principal component has an explained variance of 80% while the second principal component has 20%. Figure 3.9b shows the explained variance plot of cluster **D** along with other clusters whose first components have explained variance less than 99%.

Table 3.3 shows the top weights of the principal components of the clusters. Only the first principal components for most of the clusters are shown, as most of them have negligible explained variances for subsequent principal components. Power is the predominant variable in the principal components of all the clusters, followed by thickness of the gear-pairs in the final transmission stage. In larger clusters, module occupies higher positions compared to smaller clusters.



(a) Isomap Residual Variance for the cluster **D**. (b) PCA Explained variance for some of the clusters.

Figure 3.4: Isomap and PCA results for the clusters of fixed lay-out problem. **D** has the largest residual variance, all other clusters have negligible residual variances. The plot indicates a one dimensional manifold. Only cluster **D** has two significant principal component, all others have only one significant principal component.

3.2.2 Discussion

Apart from being a variable, power (p) is also one of the objectives of the multi-objective optimization problem, taking a different value for each pareto-optimal solution. This explains its highest weight in the principal components of all the clusters. Next we have the thicknesses of gear-pairs (t_i) with module thickness (m) occupying various positions depending on the size of the clusters. The positions of the thickness variables can be explained on the basis of input to output speed ratio for each gear-pair. These ratios are shown in table 3.4 in increasing order. The gear-pairs having smaller input to output speed ratio should have higher thickness as the pinion of these gear-pairs has to withstand higher stresses. The speed of the gear-pairs also inversely affects the stress, the higher the speed the lower will be the stress for the same power transmitted.

The positions of thickness variables in the principal components in table 3.3 have been colour coded similar to table 3.4. It can be observed that the thickness variables

A	p	t_8	t_9	m	t_7	t_4	t_5	t_6	t_1	t_3	t_2
	0.99	0.04	0.02	0.02	0.01	0.01	0.01	0.01	0.00	0.00	0.00
B	p	t_8	m	t_9	t_7	t_4	t_5	t_6	t_1	t_3	t_2
	0.99	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
C	p	t_8	t_9	m	t_7	t_4	t_5	t_6	t_1	–	–
	0.99	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0	0
D	p	t_8	t_9	t_7	m	t_4	t_5	t_6	t_1	t_3	t_2
	0.99	0.07	0.06	0.03	0.03	0.03	0.03	0.02	0.02	0.00	0.00
	t_8	t_9	t_7	t_4	t_5	t_6	t_1	p	m	t_3	t_2
	-0.64	-0.46	-0.32	-0.29	-0.26	-0.24	0.17	0.12	0.02	0.00	0.00
E	p	t_8	t_9	t_7	t_5	t_4	t_6	m	t_1	t_3	t_2
	0.99	0.07	0.05	0.03	0.02	0.02	0.02	0.01	0.01	0.00	0.00
F	p	t_8	t_9	t_7	t_4	t_6	t_5	t_1	m	t_2	t_3
	0.99	0.03	0.02	0.01	0.01	0.01	0.01	0.01	0.0	0.0	0.0
G	p	t_8	t_9	t_7	t_6	t_4	t_5	t_1	m	t_3	t_2
	0.99	0.07	0.05	0.04	0.03	0.03	0.03	0.02	0.01	0.0	0.0
H	p	t_8	t_9	t_4	t_7	t_5	t_6	t_1	m	–	–
	0.99	0.04	0.03	0.02	0.01	0.01	0.01	0.01	0.00	0	0
I	p	t_8	t_9	t_7	t_4	t_5	t_6	t_1	m	–	–
	0.98	0.10	0.07	0.05	0.04	0.04	0.04	0.03	0.01	0	0
J	p	t_8	t_9	t_7	t_4	t_6	t_5	t_1	t_3	t_2	m
	0.97	0.13	0.10	0.06	0.06	0.05	0.05	0.03	0.0	0.0	0
K	p	t_8	t_9	t_7	t_4	t_5	t_6	t_1	t_3	t_2	m
	0.85	0.33	0.25	0.17	0.14	0.13	0.13	0.09	-0.00	-0.00	0

Table 3.3: Principal component weights of each cluster in sorted order of absolute weights. Background shading indicates the input to output speeds for the gear pair. Darker cells indicate higher ratio. The gear pairs in the last transmission stage have the highest weights and those in the first transmission stage have the lowest weights for all the clusters.

Gear pair No.	1	8	9	2	4	7	5	6	3
I/O Speed ratio	0.35	0.46	0.54	0.76	0.86	0.97	0.97	1.09	1.71

Table 3.4: Gear pairs with lowest input to output speed ratios.

occur in the order of the transmission stage they are in. t_8 and t_9 are the thicknesses of the gears in the final transmission stage, t_7 in the penultimate stage followed by t_4 , t_5 , t_6 and t_1 , t_2 , t_3 in the second and first stages. The speeds would generally be the lowest in the final transmission stage, consequently, the thicknesses in these gear-pairs would have to be increased commensurate with the increasing power. As opposed to this, the shafts in the first transmission stage will have the highest speeds

and the stresses will be the lowest, hence the thickness of these gears generally stick to the lowest possible value of 0.5. Only for high power gear boxes do they show any variation, e.g. in clusters **J** and **F**, in which t_1 has higher absolute weights than module (m). It can also be noticed that gear-pairs in the same shaft are arranged in the order of increasing input to output ratio in the principal components. Although G_1 has the lowest input to output speed ratio, the position of t_1 in the principal components is after the gear-pairs in other transmission stages. This suggests that the speed of the gear pair is an important factor in stress and thickness of the gear-pair.

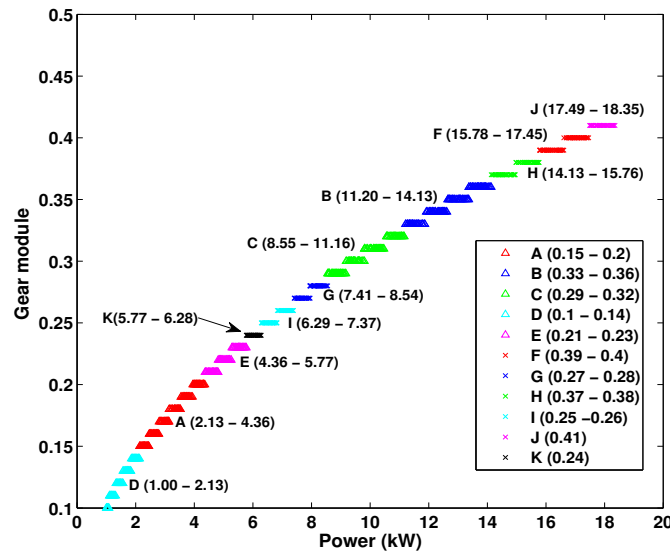


Figure 3.5: Power Vs. module characteristics of the pareto-front. Clusters occupy contiguous steps in the staircase structure.

Module weight in the principal components depends upon the size of the cluster. For clusters **A**, **B**, **C** and **D**, m occurs at positions up to fifth. Figure 3.5 shows the power Vs. module plot for the clusters. The staircase structure of this plot shows that for a higher power gearbox a larger module is required. This plot also explains the higher weights of m variable in the clusters mentioned earlier. The cluster **A** is

composed of designs with five different module widths, cluster **B** and **C** with four, and cluster **D** again with five different module widths. For these clusters, module has a good range of variation hence the higher weights in the principal components. On the other hand, clusters **J** and **K** have designs with the same module width, their principal components indicate this by having zero weights for m .

3.3 29 Variable problem

The pareto-front and the clusters obtained for the 29 variable problem are shown in the figure 3.6. As with the fixed layout problem, we keep the size and number of clusters low by choosing a higher k parameter value in our algorithm.

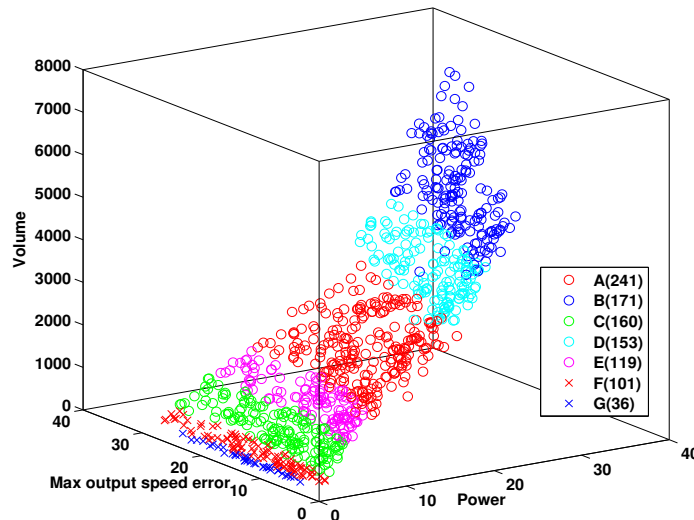


Figure 3.6: The pareto-front and the clusters for the 29-variable problem. The pareto-front manifold is clearly a two dimensional manifold in the objective space.

Figure 3.7 shows the Isomap residual variance and PCA explained variances for the pareto-front. The residual variance plot gives no certain indications about the dimensionality of the pareto-front. The largest drop in residual variance occurs for

second Isomap dimension and keeps dropping up-to the fifth dimension after which it flattens out. The explained variance also gives a similar picture; the explained variances for the first, second, third and fourth principal components are 63.6%, 11.6%, 7.6% and 7.1% respectively.

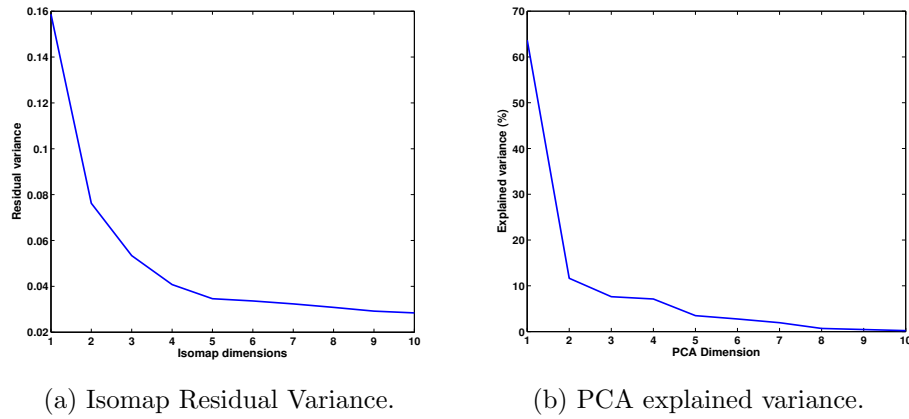


Figure 3.7: Isomap and PCA results for the 29 variable gearbox design problem pareto-front. Although the largest drop in residual variance is for two-dimensional embedding, it keeps dropping up to the fifth dimension. PCA explained variance shows four significant principal components.

Table 3.5 shows the variables with highest absolute weights of first four principal components. There are no predominant variables in the principal components, the weights are somewhat evenly distributed for the first principal component. The second principal component, however, is dominated by the p variable followed by gear teeth variables.

3.3.1 Analysis of the clusters

Cluster **A** has the largest number of optimal designs and occupies the middle portion of the power range. Cluster **B** has the highest power gearbox designs. **C** occupies the region between **A** and **B**. All other clusters have low power gearbox designs with **G** being the smallest and with the least power rating designs. Having maximum output

First PC	n_{16} 0.35	t_5 0.29	n_8 0.29	n_{10} 0.29	n_3 0.26
Second PC	p -0.81	n_2 -0.27	n_{16} -0.22	n_{18} -0.21	n_8 0.16
Third PC	n_{11} -0.43	n_9 -0.41	n_5 0.41	n_7 -0.40	n_3 0.38
Fourth PC	n_{12} 0.41	n_{10} 0.41	n_8 0.41	n_2 -0.37	n_4 -0.33

Table 3.5: Highest absolute weights of the first four principal components of the gear train design problem (29 variables). There is no dominant variable in the first principal component.

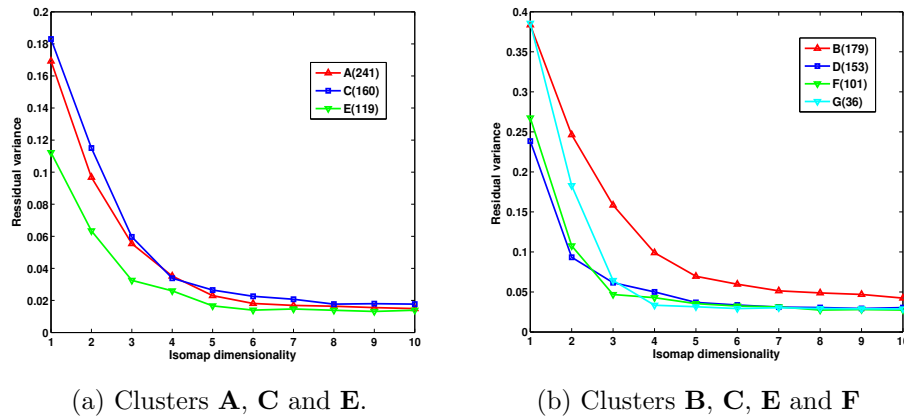


Figure 3.8: Isomap residual variances for clusters for the variable gear teeth problem. Residual variances for the clusters are similar to the whole pareto-front

speed error as an objective doesn't provide any additional useful choices to the designer, as this objective doesn't have any trade-off or conflict with the other two objectives, turning this optimization into essentially a two objective optimization.

The Isomap residual variances of the clusters shown in figure 3.8 shows the clusters have manifold dimensionality similar to the whole Pareto-front. The manifold dimensionality of the clusters maybe 1 or 2 for most of the clusters though there are no precise indications of the exact number of dimensions. The *minimization of error objective* in this case is an example of an *ill behaved* objective function as discussed in section 1.3 with reference to the chunk dimensionality conjecture, as it doesn't

monotonically increase or decrease with the n_{p_i} and n_{w_i} variables. Nevertheless, the dimensionality of the clusters are still in accordance with the conjectures claim. The explained variance plots (figure 3.9) present a picture similar to that of the Isomap residual variances. The plots show the cumulative explained variance of the principal components of the clusters. For all the clusters the cumulative explained variance goes above 95% only after the sixth principal component and the first three principal components have quite significant explained variances.

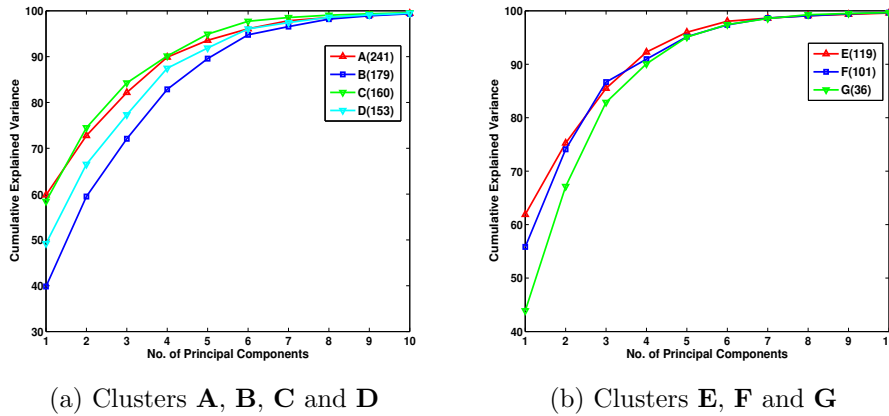


Figure 3.9: Cumulative PCA explained variance for 29 variable problem. For most clusters, first five-six principal components cover 90% of the explained variance.

In all the principal components number of teeth variables have the highest weights. Unlike in the fixed layout gearbox design problem, power is not the dominant variable in the principal components. The t_i variables have weights lower than most of the number of the n_i variables, except n_{15} , n_{16} , n_{17} and n_{18} . Module is among the least important variables in all the principal components.

We have already explored the significance of the thickness variables and module thickness in the pareto-front in the previous section. For this reason, we will concentrate on the parts that the n_i variables play in the pareto-front clusters. Table 3.6 lists the n_i variables in the sorted order of their absolute weights in the principal

components. The cells are shaded according to the transmission stage the variable belongs to. The number of teeth variables of the second transmission stage (n_7 to n_{12}) show the most amount of variability while the last transmission stage has the most stable configuration showing the least amount of variation in all the clusters.

A	First PC	n_8	n_9	n_{12}	n_6	n_7	n_{11}	...	n_{16}	n_{18}	n_{15}	n_{17}
	Second PC	n_{11}	n_{13}	n_{10}	n_{12}	n_5	n_6	...	n_{17}	n_{16}	n_{15}	n_{18}
B	First PC	n_{12}	n_8	n_{11}	n_6	n_5	n_9	...	n_{17}	n_{15}	n_{16}	n_{18}
	Second PC	n_{11}	n_{10}	n_8	n_{12}	n_6	n_{13}	...	n_{15}	n_{17}	n_{16}	n_{18}
C	First PC	n_{11}	n_7	n_{10}	n_6	n_9	n_5	...	n_{17}	n_{18}	n_{15}	n_{16}
	Second PC	n_{12}	n_6	n_8	n_{11}	n_9	n_{10}	...	n_{15}	n_{16}	n_{18}	n_{17}
D	First PC	n_9	n_8	n_{10}	n_5	n_{12}	n_7	...	n_{15}	n_{16}	n_{17}	n_{18}
	Second PC	n_{10}	n_{12}	n_{11}	n_7	n_6	n_8	...	n_{17}	n_{15}	n_{18}	n_{16}
E	First PC	n_8	n_9	n_{12}	n_6	n_{11}	n_{13}	...	n_{18}	n_{17}	n_{15}	n_{16}
	Second PC	n_9	n_8	n_{12}	n_7	n_{10}	n_4	...	n_{16}	n_{15}	n_{17}	n_{18}
F	First PC	n_{11}	n_9	n_6	n_8	n_5	n_7	...	n_{16}	n_{15}	n_{18}	n_{17}
	Second PC	n_8	n_6	n_7	n_9	n_{11}	n_5	...	n_{16}	n_{18}	n_{17}	n_{15}
G	First PC	n_{11}	n_{13}	n_6	n_7	n_9	n_8	...	n_{17}	n_{15}	n_{16}	n_{18}
	Second PC	n_{12}	n_4	n_8	n_{10}	n_9	n_{11}	...	n_{17}	n_{16}	n_{15}	n_{18}

Table 3.6: Ranking of gear teeth variables on the basis of absolute weights in principal components. The gear pairs of the last transmission stage occupy the last positions indicating they are the least varying.

3.3.2 Discussion

Although power is still an objective in the multi-objective optimization, since there are three objectives (*volume* and *maximum error in output speeds* being the other two), a number of designs with same power rating are possible, hence power is not a dominant variable in the principal components. As observed earlier, having error as a third objective isn't useful in the design process, as having higher error apparently neither reduces the volume significantly nor improves the power characteristics of the designs.

The final transmission stage has to withstand the highest stresses, so the module

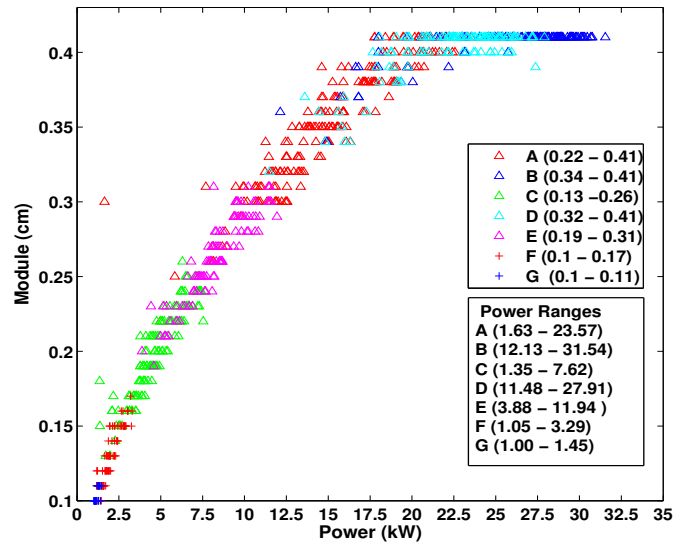


Figure 3.10: Power Vs. module characteristics of the pareto-front for 29 variable problem. Clusters show intermingling unlike for the fixed layout problem.

choice for the gear-box has to be able to withstand the stresses in the final transmission stage. The same module is used in the gears of all the transmission stages. The number of teeth has to be kept minimum possible as it also affects gearbox volume directly. Also there is a limit on the maximum transmission ratio. All these trade-offs and limitations lead to limited choices for the number of teeth variables in the final transmission stage. On the other hand, in the second transmission stage ($n_7 - n_{12}$) has to withstand smaller stresses than the last stage, it can meet the same stress requirements by increasing the number of teeth and decreasing the gear pair thicknesses while maintaining the required transmission ratios. This explains the positions of the number of teeth variables in the last and second transmission stage.

The module to power characteristics are similar to the fixed layout case, as shown in figure 3.10. Unlike in the previous case, designs in different clusters can have the same module width.

Chapter 4

More Design Problems

In this chapter we present the results of our chunking and analysis methodology on two problems. The first problem is the multiple disk clutch brake design problem, a discrete optimization problem with five variables and two objectives. The other is the welded beam design problem which is a continuous decision space problem with four variables and two objectives.

4.1 Problem Introduction

4.1.1 Multiple-disk clutch brake design problem

We have adopted the problem specification used in [Deb and Srinivasan, 2006]. The schematics of a multiple-disk clutch brake are shown in figure 4.1. Two conflicting objectives are considered in this optimization problem:

1. minimization of mass, and,
2. minimization of stopping time.

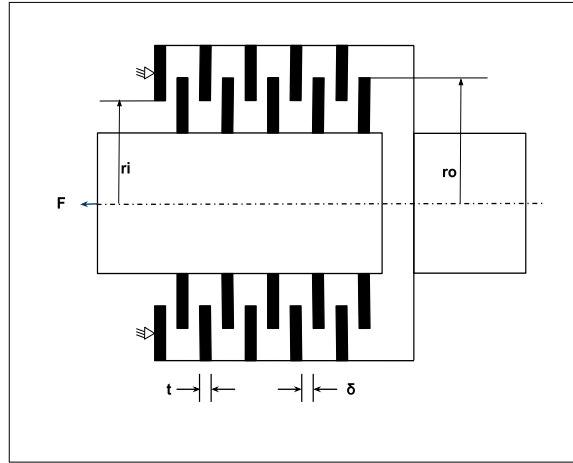


Figure 4.1: Schematics of a multiple disk clutch brake.

This optimization problem is defined on five decision variables $\vec{x} = (r_i, r_o, t, F, Z)$, where:

1. $r_i \in [60, 80]$ (in steps of one) is the inner radius in mm,
2. $r_o \in [91, 110]$ (in steps of one) is the outer radius in mm,
3. $t \in [1, 3]$ (in steps of 0.5) is the thickness of disks in mm,
4. $F \in [600, 1000]$ (in steps of 10) is the actuating force in N and
5. $Z \in [2, 10]$ (in steps of one) is the number of friction surfaces or disks.

The complete optimization problem is formulated as follows:

$$\text{Maximize } f_1(\vec{x}) = \pi(x_2^2 - x_1^2)x_3(x_5 + 1)\rho, \quad (4.1)$$

$$\text{Minimize } f_2(\vec{x}) = T = \frac{I_z\omega}{M_h + M_f}, \quad (4.2)$$

Subject to,

$$g1(\vec{x}) = x_2 - x_1 - \Delta R \geq 0, \quad (4.3)$$

$$g2(\vec{x}) = L_{max} - (x_5 + 1)(x_3 + \delta) \geq 0 \quad (4.4)$$

$$g3(\vec{x}) = p_{max} - p_{rz} \geq 0 \quad (4.5)$$

$$g4(\vec{x}) = p_{max}V_{sr,max} - p_{rz}V_{sr} \geq 0, \quad (4.6)$$

$$g5(\vec{x}) = V_{sr,max} - V_{sr} \geq 0, \quad (4.7)$$

$$g6(\vec{x}) = M_h - sM_s \geq 0, \quad (4.8)$$

$$g7(\vec{x}) = T \geq 0, \quad (4.9)$$

$$g8(\vec{x}) = T_{max} - T \geq 0 \quad (4.10)$$

Where,

$$\bullet M_h = \frac{2}{3} \mu x_4 x_5 \frac{x_2^3 - x_1^3}{x_2^2 - x_1^2} \text{ N.mm},$$

$$\bullet \omega = \frac{\pi n}{30} \text{ rad/s},$$

$$\bullet A = \pi(x_2^2 - x_1^2) \text{ mm}^2,$$

$$\bullet p_{rz} = \frac{x_4}{A} \text{ N/mm}^2,$$

$$\bullet V_{sr} = \frac{\pi R_{sr} n}{30} \text{ mm/s},$$

$$\bullet R_{sr} = \frac{2}{3} \frac{x_2^3 - x_1^3}{x_2^2 - x_1^2} \text{ mm},$$

$$\bullet \Delta R = 20 \text{ mm},$$

$$\bullet L_{max} = 30 \text{ mm},$$

$$\bullet \mu = 0.5,$$

$$\bullet p_{max} = 1 \text{ MPa},$$

$$\bullet \rho = 0.0000078 \text{ kg/mm}^3,$$

$$\bullet V_{sr,max} = 10 \text{ m/s},$$

$$\bullet s = 1.5,$$

$$\bullet T_{max} = 15 \text{ s},$$

$$\bullet n = 250 \text{ rpm},$$

$$\bullet M_s = 40 \text{ Nm},$$

$$\bullet M_f = 3 \text{ Nm},$$

$$\bullet I_z = 55 \text{ kg.m}^2,$$

$$\bullet \delta = 0.5 \text{ mm},$$

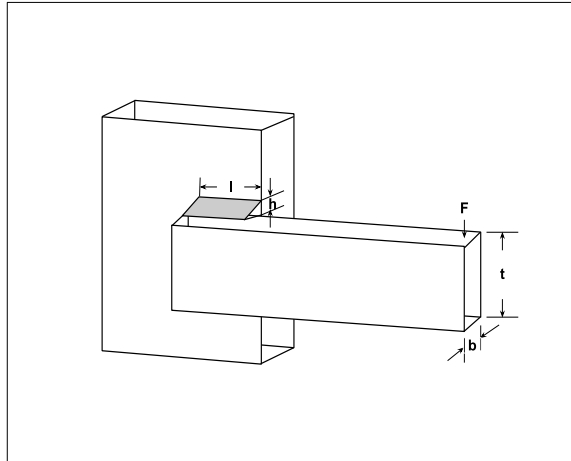


Figure 4.2: Schematics of a welded beam.

4.1.2 The welded beam design problem

Welded beam design problem is a well known Multi-objective optimization problem. Here we adopt the formulation given in [Bandaru and Deb, 2010]. The objective of this problem is to minimize the cost and end deflection of a beam welded at one end and supporting a load $F = 6000$ lb at the other. The overhang portion of the beam has a length of 14 in. There are four design variables:

1. $b \in [0.125, 5]$ thickness of the beam,
2. $t \in [0.1, 10]$ width of the beam,
3. $l \in [0.1, 10]$ length of the weld and
4. $h \in [0.125, 5]$ thickness of the weld.

The complete multi-objective formulation is as follows:

$$\text{Minimize } f_1(\vec{x}) = C = 1.1047h^2l + 0.04811tb(14.0 + l) \quad (4.11)$$

$$\text{Minimize } f_2(\vec{x}) = D = \frac{2.1952}{t^3b} \quad (4.12)$$

Subject to:

$$g_1(\vec{x}) = 13600 - \tau(\vec{x}) \geq 0, \quad (4.13)$$

$$g_2(\vec{x}) = 30000 - \sigma(\vec{x}) \geq 0, \quad (4.14)$$

$$g_3(\vec{x}) = b - h \geq 0, \quad (4.15)$$

$$g_4(\vec{x}) = P_c(\vec{x}) - 6000 \geq 0 \quad (4.16)$$

Where,

- $\tau(\vec{x}) = \sqrt{\frac{(\tau')^2 + (\tau'')^2 + (l\tau'\tau'')}{\sqrt{0.25(l^2 + (h+t)^2)}}$,
- $\tau' = \frac{6000}{\sqrt{2hl}}$,
- $\tau'' = \frac{6000(14 + 0.5l)\sqrt{0.25(l^2 + (h+t)^2)}}{2[0.707hl(\frac{l^2}{12} + 0.25(h+t)^2)]}$,
- $\sigma(\vec{x}) = \frac{504000}{t^2b}$,
- $P_c(\vec{x}) = 64746.022(1 - 0.0282346t)tb^3$.

4.2 Analysis of the clutch brake design problem

Figure 4.3 shows the pareto-front obtained after running local search on NSGA-II output. NSGA-II was run for 300 generations with a population size of 5000 and yielded 125 near optimal solutions. The probability for crossover and mutation were set to 0.79 and 0.5 respectively. Running local search gives a pareto-front consisting of 95 optimal solutions.

4.2.1 Isomap and PCA analysis of the pareto-front

Figure 4.4a and 4.4b show the results of running Isomap and PCA on the pareto-front of the clutch brake design problem. The drop in the residual variance curve

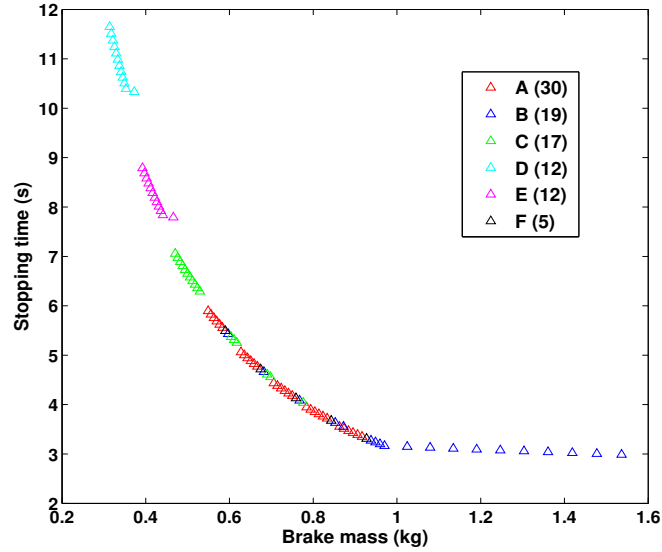


Figure 4.3: Pareto-front and the clusters of the multiple-disk clutch brake design problem. Clusters obtained with $k = 1.9$

for the second Isomap dimension indicates a manifold dimensionality of one. The PCA explained variance, however, shows two principal components with significant explained variances, the first with 80% explained variance and the second with 17%.

Table 4.2 shows the first two significant components of the pareto-front. Outer radius (r_o) has the highest weight in the first principal component followed by inner radius (r_i). The number of surfaces also has some variation. In the second principal component number of surfaces (Z) is the dominant variable followed by inner radius with a small weight. The variables thickness (t) and actuating force (F) have zero weights in both the significant principal components indicating they have the same values for all the optimal solutions. The thickness and actuating force are fixed at the minimum and maximum possible values of 1 and 1000 respectively for all the optimal solutions.

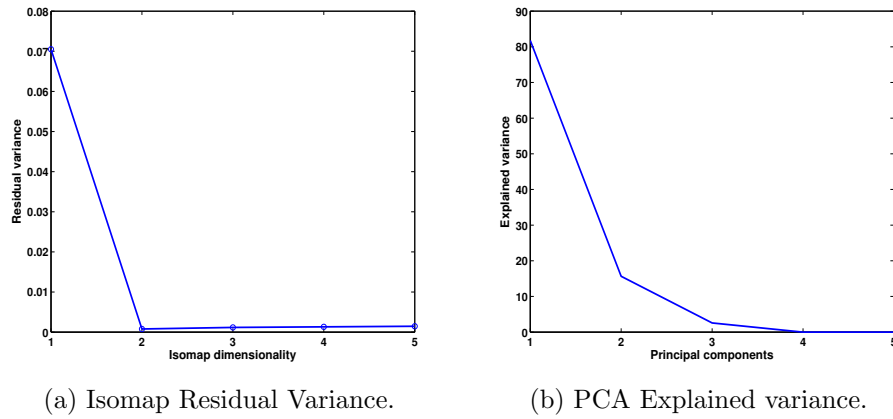


Figure 4.4: Isomap and PCA results for the pareto-front of multiple-disk clutch brake problem pareto-front. The manifold dimensionality is one as indicated in the Isomap residual variance plot but the PCA explained variance shows two significant principal components.

	r_i	r_o	t	F	Z
First PC	0.578	0.806	0	0	0.119
Second PC	-0.207	0.004	0	0	0.978

Table 4.1: First two principal components of the clutch brake design problem pareto-front.

4.2.2 Clustering analysis

The pareto-front of this problem has a small number (95) of optimal solutions. The k parameter of the clustering algorithm had to be adjusted accordingly to obtain a proper distribution of points in clusters. Six clusters were obtained with $k = 1.9$. The clusters are shown in figure 4.3.

Cluster **A** is the largest cluster with 37 points. It has the mid-range designs of the clutch brake with stopping times ranging from 3.2 seconds to 6 seconds. Cluster **B** is composed of the heaviest and most powerful brake designs, apart from some mid-range designs. It's the only cluster having designs with weights in excess of 1 kg. Below 1 kg designs follow a parabolic time vs. weight characteristics, but above 1 kg designs show a linear relationship, an increase in weight yielding very little

benefit in stopping time. Clusters **D** and **E** have the low end designs with stopping times in excess of 7.5 s. These are the only clusters that appear as single contiguous chunks in the pareto-front. Clusters **A**, **B**, **C** and **E** have designs in overlapping objective ranges.

Isomap residual variances of the clusters (figure 4.5a) indicate one dimensional manifolds for the clusters, similar to the whole pareto-front. This verifies the claims of the *chunk dimensionality conjecture* (section 1.3). Clusters **B-C** and **D-E** have exactly same residual variances. Cluster **A** has the highest error in reverse mapping the one dimensional Isomap embedding. Clusters **D** and **E** have negligible residual variances. PCA explained variances for the clusters (figure 4.5b) show a single significant principal component for all the clusters except cluster **A**.

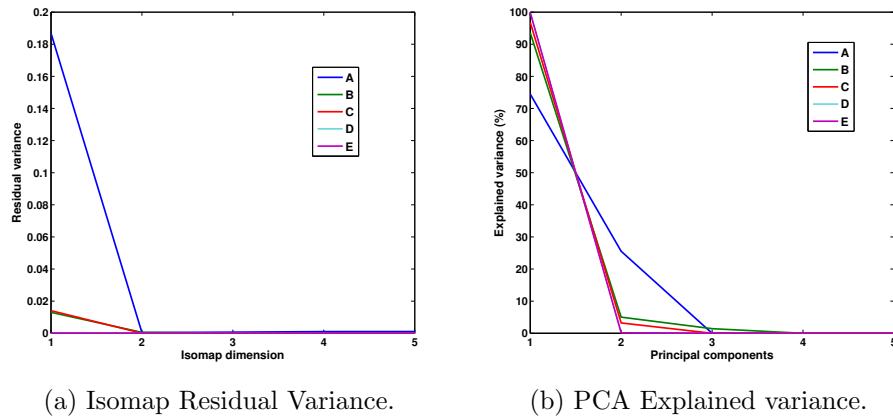


Figure 4.5: Isomap and PCA results for the clusters of multiple-disk clutch brake problem. The manifold dimensionality of all the clusters is one, though cluster **A** has linear dimensionality of two indicated by two significant principal components in its explained variance plot.

Table 4.2 shows the principal components of the clusters. For most of the clusters, the radius variables have similar weights. In cluster **B** the outer radius has higher weight than the inner radius. The second principal component of cluster **A** has entirely composed of a single variable Z only. Clusters **D** and **E** have exactly the

	r_i	r_o	t	F	Z
A	0.707	0.707	0	0	0
	0	0	0	0	1
B	0.236	0.960	0	0	0.146
C	0.703	0.703	0	0	0.097
D	0.694	0.719	0	0	0
E	0.694	0.719	0	0	0
F	0	0	0	0	1

Table 4.2: Significant principal components of the clutch brake design problem clusters. t and F have no weights in any principal component. The inner and outer radii (r_i and r_o) show the highest variation for each cluster. The second component is parallel to the *number of friction surfaces* (Z) dimension.

same components. The cluster **F** has variable Z as its single Principal component.

4.2.3 Discussion

The most important fact that comes out of this analysis is that all optimal clutch brakes should be designed with disks with smallest possible thickness. Secondly, a clutch brake must be designed for maximum actuating force only, since the stopping time is inversely proportional to the actuating force. The stopping time is also dependent on the friction area as is evident from the figure 4.6. A higher outer radius and a larger number of friction surfaces (no. of disks) allow for a larger contact area, but increasing the number of disks also increases the weight of the design. This trade-off is visible in all the clusters. The fact that radius variables show greater variation across all clusters suggests that to reduce the stopping time increasing the radial dimension is the more economical option.

The two dimensional nature of cluster **A** can be explained by this plot. **A** has designs with small variation in radial dimensions except in very heavy clutch brakes, those occupying the linear region in plot 4.3. For the linear region designs, the number of disks is maximum and the only way to design more powerful brakes is by

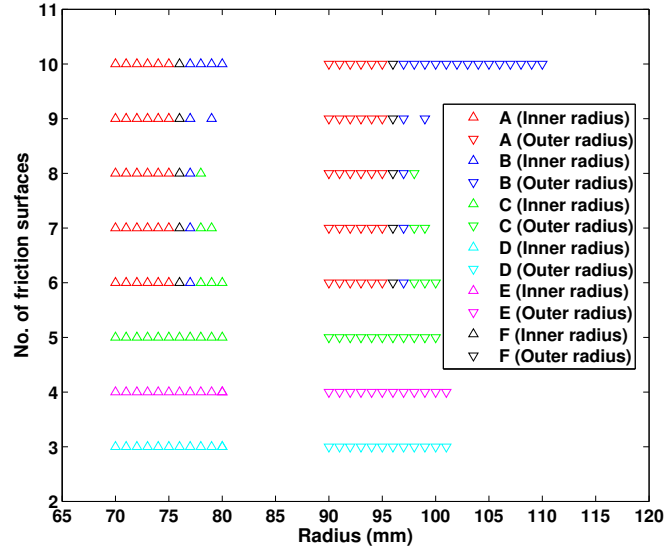


Figure 4.6: Radius Vs. no. of friction surface plot for pareto-front clusters. Clusters **A**, **B**, **C** and **F** are distributed in four lines each while **D**, **E** and **F** are composed of single lines.

increasing the outer radius to increase the friction area. If the clustering had been done in the design variables space only we would have clusters on the basis of the number of contact surfaces only. Clustering in the combined objective-decision variable space gives more meaningful clusters, which have functionally similar designs.

4.3 Analysis of welded beam design problem

The NSGA-II run for the welded beam design problem yielded 500 solutions with a population size of 500 running for 200 hundred generations. The crossover and mutation probability were set to 0.73 and 0.53. After local search the pareto-front obtained had 491 optimal solutions. Figure 4.7 shows the final pareto-front.

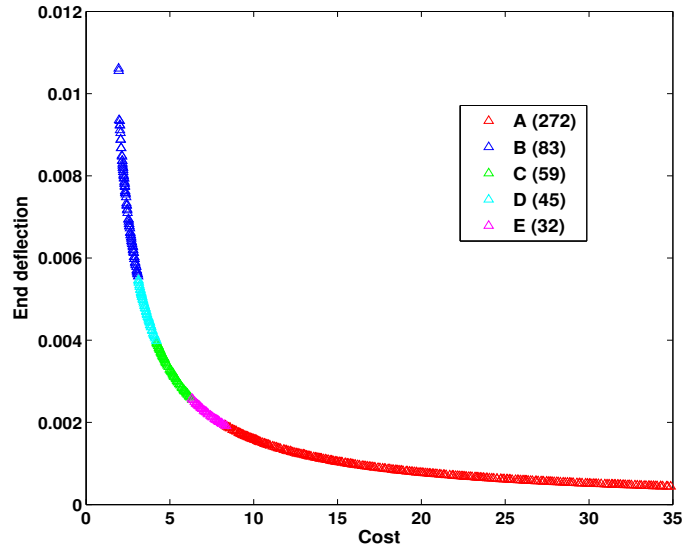


Figure 4.7: Pareto-front and clusters of the welded beam design problem. Obtained with $k = 2$. The clusters in the extremes are the largest while those around the knee of the curve are relatively smaller.

4.4 Isomap and PCA analysis of the pareto-front

The residual variance plot shown in figure 4.8a shows the largest drop for dimension 2. There is some further drop going from second to third dimensional embedding. This indicates a manifold dimensionality of one. The PCA explained variance (figure 4.8b) also shows only one significant principal component with an explained variance of 93%. Table 4.3 shows the first principal component of the pareto-front. The beam thickness (b) is the design variable with the highest weight in the principal component. Weld thickness (h) and length of the weld (l) have similar weights, albeit that of l being in the reverse direction. Width of the beam (t) is the least varying variable of all.

	b	t	l	h
First PC	0.903	0.001	-0.278	0.325

Table 4.3: Principal component of the clutch brake design problem pareto-front

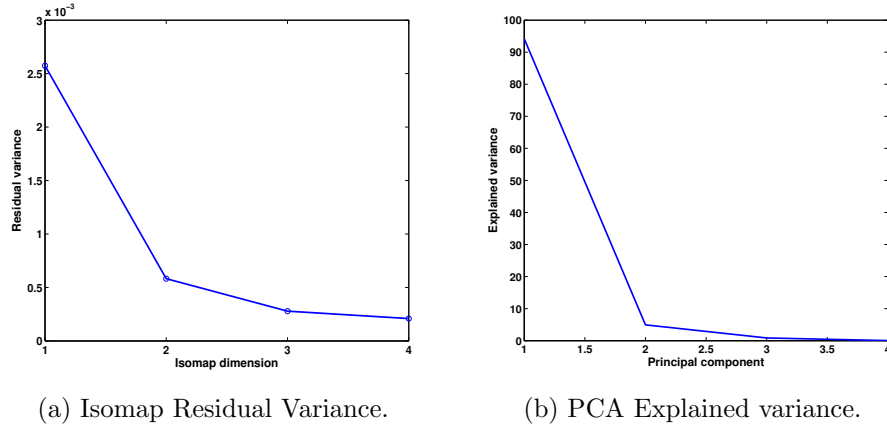


Figure 4.8: Isomap and PCA results for the welded beam design problem pareto-front.

4.4.1 Clustering analysis

A large number of specific clusters could be obtained by suitably setting the parameters of the algorithm, but we analyze only 5 clusters for the sake of simplicity. The clusters are shown in figure 4.7.

Clusters **A** and **B** have the largest number of optimal solutions. They occupy the linear extremes of the pareto-front curve. Three clusters are concentrated around the knee of the curve. Cluster **A** has the costliest and sturdiest designs, while cluster **B** has the cheapest and the weakest beams. Clusters **C**, **D** and **E** have beams with mid range end deflection and cost.

The Isomap residual variances for all the clusters suggest one dimensional manifolds clusters as all the curves have the largest drop for second dimension. Clusters **D** and **C** have relatively higher residual variances than other clusters. The PCA explained variance for these clusters also show two significant components. The explained variance plots for clusters **A** and **B** are almost the same resulting in overlapping lines in the plot. The cluster **E** also has two significant components.

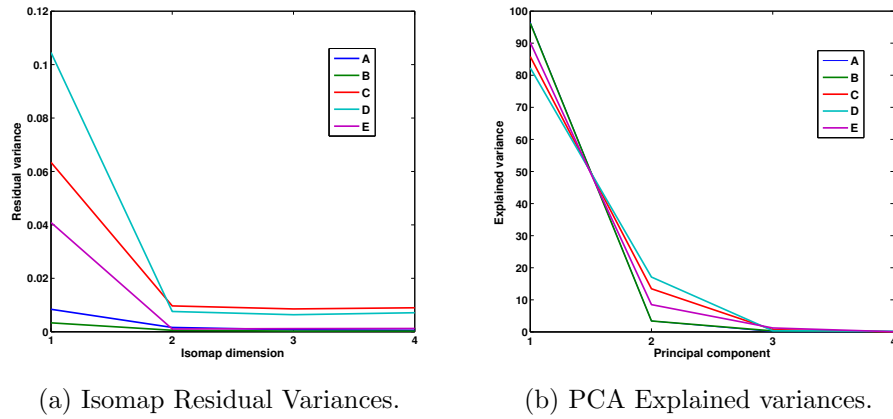


Figure 4.9: Isomap and PCA results for the clusters of welded beam design problem. All the clusters are one dimensional manifolds but clusters **C** and **D** have two linear dimensions as they have two significant components.

Thickness of the beam (b) is the predominant variable in the principal component of the cluster **A** (table 4.4) while length of the weld (l) is the predominant variable in cluster **B**. These two are at the two extremes of the pareto-front, the former having designs with least deflection and highest cost and the latter having cheapest designs with large end deflections. The clusters around the knee of the pareto-front curve also have their two significant components in the $b - l$ plane with first component leaning more towards the l direction and second component leaning towards b . For all the clusters width of the beam (t) has the least variation. For all the optimal designs, the width of the beam is approximately maximum possible value of 10.

4.4.2 Discussion

Width of the beam (t) variable has the least variation for all the clusters. All the optimal designs use the maximum possible value of the width beam. This suggests that a welded beam should be as wide as possible, other design variables should be adjusted to obtain suitable designs. In cluster **A** of the sturdiest designs, designs

	b	t	l	h
A	0.965	0	-0.06	0.249
B	0.126	0.068	-0.982	0.115
C	0.435	0	-0.755	0.488
	0.897	0.006	0.404	-0.174
D	0.309	0	-0.883	0.351
	0.949	0.011	0.307	-0.06
E	0.381	0.002	-0.616	0.688
	0.924	0.019	0.256	-0.282

Table 4.4: Significant principal components of the pareto-front clusters of the welded beam design problem. The one dimensional clusters have b and l as the predominant variables. Two dimensional clusters also have their significant components in the b - l plane. t is the least varying variable in all the clusters.

tend to vary in the thickness of the beam the most with other variables limited to small ranges of variation. The maximum value of l and h are 1.39 and 2.49 which are significantly less than their maximum possible of 10 and 5 respectively. This suggests that increasing the size of the weld in l and h beyond certain limit doesn't help in improving the sturdiness much, though increasing thickness of the beam may improve the end deflection. In cluster **B** the cost is minimum though the designs have larger end deflection than in other clusters. In this cluster, l is the variable that varies the most, with other variables showing less variation. l is increased or decreased to increase or decrease the cost.

Chapter 5

Conclusion

We have proposed a clustering and dimensionality reduction approach to chunking in design tasks. A multi-objective optimization procedure is used to first obtain a set of good designs, then clustering on the combined objective-decision variable space is used to group the designs according to their functional characteristics. Linear and non-linear dimensionality reduction techniques are used to bring out the design implications of these clusters and capture their semantics. We have applied this procedure to a number of practical objective optimization problems and shown that majority of the clusters obtained are one dimensional. In the examples considered pareto frontiers that are two or higher dimensional manifolds, the clustering neatly separated the higher dimensional patches of the manifold into separate clusters.

This procedure could not only be used to learn design symbols in the cognitive sense as discussed in [Mukerjee and Dabbeeru, 2009], but also to specialize them further. For example, in BDCPMM design cluster **A** and **B** would correspond to what the expert designer calls “low end designs” (section 2.7) and the cluster **C** to “high end designs”. This would provide a richer initial vocabulary to the system

with some interrelations among symbols already established.

The analysis of any decision task involves moving through several layers of complexity. This suggests how a system can enrich already learnt symbols through experience. Initially, the system could be presented with a minimal multi-objective formulation (two objectives) of a design task. When the system has learnt some initial symbols, a more detailed formulation with possibly more objectives and a larger number of decision variables could be presented to the system to gain “experience”. The initial minimal formulation would help establish the most pertinent principles, and the gradual increase in complexity of the design task would help to bring out the subtle and implicit relationships among the chunks. For example in the gearbox design problem, the analysis of the fixed layout version of the problem established the fact that thickness of the gear pairs in the last transmission stage should go higher in accordance with the higher power requirements, and the analysis of the complete problem brought to the front some ideal combinations for no. of teeth in the final transmission stage. However, the specification of the minimal initial formulation, and the appropriate increase in complexity after each learning stage would depend on the design task at hand.

We have shown the application the procedure to problems in design domain, but such an approach to chunking is possible in any domain where the task can be expressed as a multi-objective optimization. For example, in robot path planning [Vadakkepat et al., 2002], many potential field functions are optimized to calculate the optimal path for the robot. However, optimization algorithms are complex and time consuming hence not suited for real time application, but if a robot were to learn from the optimizations as it calculates the optimal navigation path for a given

situation, it could use the readily available learnt information when faced with a similar scenario next time. With increasing experience, such a system may be able to learn chunks such as “door”, “connecting passage”, “detours” as components of high-level abstraction for important tasks.

Finally we return to the *chunk dimensionality conjecture*. In the five examples considered the dimension of the chunk manifold is comparable to the number of objectives. However as empirical validation, the sample set is very small and clearly much work needs to be done to come up with a theoretical basis for the conjecture.

Clearly the ramifications of such a conjecture are very high, especially in the attempt to create machine models of expert behaviour. In this thesis we have presented evidence that suggests that for many practical problems such a conjecture may hold. This work may thus open the door for much greater testing and validation in order to identify precisely the problem domains where it may apply.

Bibliography

- Anderson, J. (1996). *The architecture of cognition*. Lawrence Erlbaum.
- Bandaru, S. and Deb, K. (2010). Automating discovery of innovative design principles through optimization. Technical Report 2010001, Kanpur Genetic Algorithms Laboratory, Indian Institute of Technology, Kanpur.
- Bransford, J., Brown, A., Cocking, R., et al. (2000). *How people learn*. National Academy Press Washington, DC.
- Chase, W. and Simon, H. (1973). Perception in chess. *Cognitive psychology*, 4(1):55–81.
- Chidambaram, B. and Agogino, A. (1999). Catalog-based customization. In *Proceedings of DETC*, volume 99, pages 12–16. Citeseer.
- De Groot, A. (1965). *Thought and choice in chess*. Amsterdam University Press.
- Deb, K. (2001). *Multi-objective optimization using evolutionary algorithms*. Wiley.
- Deb, K. and Jain, S. (2003). Multi-speed gearbox design using multi-objective evolutionary algorithms. *Journal of Mechanical Design*, 125:609.
- Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T. (2002). A fast and elitist multi-objective genetic algorithm: Nsga-ii. *Evolutionary Computation, IEEE Transactions on*, 6(2):182–197.
- Deb, K. and Sindhya, K. (2008). Deciphering innovative principles for optimal electric brushless dc permanent magnet motor design. In *IEEE Congress on Evolutionary Computation, 2008. CEC 2008. (IEEE World Congress on Computational Intelligence)*, pages 2283–2290. IEEE.

- Deb, K. and Srinivasan, A. (2006). Innovization: Innovating design principles through optimization. In *Proceedings of the 8th annual conference on Genetic and evolutionary computation*, page 1636. ACM.
- Domeshek, E. and Kolodner, J. (1991). Toward a case-based aid for conceptual design. *International Journal of Expert Systems*, 4(2):201–220.
- Egan, D. and Schwartz, B. (1979). Chunking in recall of symbolic drawings. *Memory & Cognition*, 7(2):149.
- Ehrlich, K. and Soloway, E. (1984). An empirical investigation of the tacit plan knowledge in programming. In *Human factors in computer systems*, pages 113–133. Intellect Books.
- Elman, J. (2005). Connectionist models of cognitive development: where next? *Trends in cognitive sciences*, 9(3):111–117.
- Engle, R. and Bukstel, L. (1978). Memory processes among bridge players of differing expertise. *The American Journal of Psychology*, 91(4):673–689.
- Ericsson, K. (2002). Attaining excellence through deliberate practice: Insights from the study of expert performance. *The pursuit of excellence through education*, pages 21–55.
- Feigenbaum, E. (1961). The simulation of verbal learning behavior. In *Western Joint IRE-AIEE-ACM Computer Conference*, pages 121–132. ACM.
- Glaser, R. (1999). Expert knowledge and processes of thinking. *Learning and knowledge*, pages 88–102.
- Gobet, F. (1993). A computer model of chess memory. In *Proceedings of the fifteenth annual conference of the Cognitive Science Society: June 18 to 21, 1993, Institute of Cognitive Science, University of Colorado, Boulder*, page 463. Lawrence Erlbaum.
- Goel, V. and Pirolli, P. (1992). The structure of design problem spaces. *Cognitive Science*, 16(3):395–429.

- Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, 24(6):417–441.
- Jain, P. and Agogino, A. M. (1990). Theory of design: An optimization perspective. *Mechanism and Machine Theory*, 25(3):287 – 303.
- Jolliffe, I. (2002). *Principal Component Analysis*. John Wiley & Sons, Ltd, second edition.
- Martin, S. and Backer, A. (2005). Estimating manifold dimension by inversion error. In *SAC '05: Proceedings of the 2005 ACM symposium on Applied computing*.
- Moss, J., Cagan, J., and Kotovsky, K. (2004). Learning from design experience in an agent-based design system. *Research in Engineering Design*, 15(2):77–92.
- Mukerjee, A. and Dabbeeru, M. M. (2009). The birth of symbols in design. In *ASME 2009 International Design Engineering Technical Conferences and Computers and Information Engineering Conference August-September*.
- Saul, L. and Roweis, S. (2003). Think globally, fit locally: unsupervised learning of low dimensional manifolds. *The Journal of Machine Learning Research*, 4:119–155.
- Schaffer, J. (1985). Multiple objective optimization with vector evaluated genetic algorithms. In *Proceedings of the 1st International Conference on Genetic Algorithms*, pages 93–100. L. Erlbaum Associates Inc., L. Erlbaum Associates Inc.
- Schön, D. (1988). Designing: Rules, types and words. *Design studies*, 9(3):181–190.
- Srinivas, N. and Deb, K. (1994). Multiobjective optimization using nondominated sorting in genetic algorithms. *Evolutionary Computation*, 2(3):221–248.
- Sycara, K., Chandra, D., Guttal, R., Koning, J., and Narasimhan, S. (1991). Cadet: a case-based synthesis tool for engineering design. *International Journal of Expert Systems*, 4(2):157–188.
- Tenenbaum, J., Silva, V., and Langford, J. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323.

Vadakkepat, P., Tan, K., and Ming-Liang, W. (2002). Evolutionary artificial potential fields and their application in real time robot path planning. In *Proceedings of the 2000 Congress on Evolutionary Computation, 2000.*, volume 1, pages 256–263. IEEE.

Van Der Maaten, L., Postma, E., and Van Den Herik, H. (2007). Dimensionality reduction: A comparative review. Published online on Citeseer.