MDO Group: HDR Norms and Expectations

Tapabrata Ray

Disclaimer: These are my personal views and do not reflect expectations of all research groups at UNSW.

Research Thesis

Your thesis is evaluated. So its extremely important to know what is the minimum requirement that leads to a thesis.

3.1. Minimum requirements

(a) An independent introduction that contextualises the research in relation to the current knowledge in the field.

(b) Thesis chapters are in a logical and cogent sequence presenting an argument that supports the main findings of the thesis.

(c) An independent discussion that integrates the significant findings of the thesis.

(d) A conclusion that summarises the findings and articulates clearly the new contribution to knowledge in the discipline.

(e) A candidate may submit work as part of the thesis that has been published or accepted for publication or manuscripts submitted for publication that contribute directly to their argument and support their findings (see Section 5.2). A thesis with publications incorporated must also meet the above minimum requirements.

(f) In addition, there are alternative formats for the HDR thesis that may be suitable for some disciplines and/or for some candidates, such as the incorporation of portfolios of creative works and/or exhibitions of practice-based research. Faculty HDCs will provide guidance on discipline specific requirements.

This is what your supervisor needs to agree to when he/she recommends the thesis for examination.

Research Thesis

This is what the examiner is asked to assess.

The examiners are asked to examine the thesis against the following criteria:

- Does the candidate demonstrate a significant and original contribution to knowledge (relative to the level of the degree being sought)?
- Does the candidate engage with the literature and the work of others?
- Does the candidate show an advanced knowledge of research principles and methods related to the applicable discipline?
- Is there a clear and discernible lucidity in the presented research, its arguments and conclusions?
- Is the thesis clearly, accurately and logically written?

This is what the primary supervisor will have to endorse.

(b) The primary supervisor must use iThenticate to confirm that the final work is free of plagiarism and suitable for examination. This will be documented on the supervisor's certificate before submission of the thesis.

(e) It is expected that published work included in the thesis has been published in high-quality peer-reviewed journals or other locations that are discipline appropriate.

Research Practice

EXPECTATIONS

The expectations vary across groups and so does across fields. Three publications in top tier journals might be great in one and poor in another field. If unsure, check with supervisor.

There has to be clear demonstration of three SIGNIFICANT and NOVEL pieces of work in a PhD thesis.

Understanding of what is NOVEL and what is SIGNIFICANT is often the thing that a student struggles to understand.

Literature Review

Gaps and Contributions

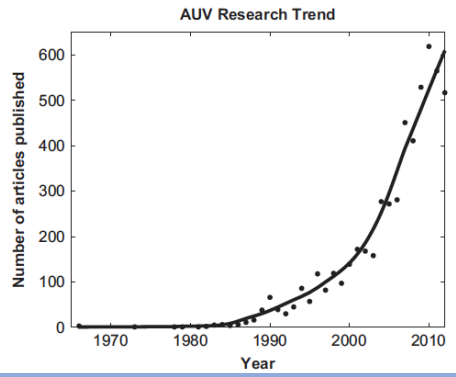
Results and Analysis

Literature Review

Have a purpose, drive the reader towards your contributions, critically assess past contributions to highlight why these aspects are important.

Use keywords and search across databases e.g. SCOPUS, Google Scholar or specific journals.

A survey was done based on a search made on *Scopus* that covers nearly 18,000 titles from more than 5000 international publishers, including coverage of 16,500 peer-reviewed journals in the scientific, technical, medical and social sciences field, with the keywords *AUV* or *Autonomous Underwater Vehicle*. From the survey, 5393 most relevant articles that were published before June 2013 have been used to conduct a research trend analysis.



An example to highlight increased activity

Table 1: Summary of literature on mean compliance minimization.

Paper	Uncertainty type	Summary
Guest and Igusa (2008)	Load components and load locations as distributions with arbitrary covariance	General scheme for handling uncertain loads and load location. Derived an approximately equivalent load distribution to result in the same compliance as the random node location.
Dunning et al. (2011)	Concentrated load magnitude and direction as distributions	Derived efficient formulations for the mean compliance, where the number of linear systems to be solved scales linearly with the number of independent random variables. Assumes that the distribution of the forces' magnitudes and rotations are known and independent.
Zhao and Wang (2014b)	Concentrated load magnitude and direction and distributed load as distributions or random fields	Requires fewer linear system solves per independent random variable compared to the approach by Dunning et al. (2011). Demonstrated how the same approach can handle distributed load uncertainty, modeled as a stochastic field, using K-L expansion.
Zhang et al. (2017)	Finite set of load scenarios	Proposed an efficient way to compute the mean compliance. Developed a randomized algorithm inspired by Hutchinson's trace estimator Hutchinson (1990) to minimize the mean compliance leading to significant computational savings compared to the naive approach.
Liu and Wen (2018)	Fuzzy load	Modeled the load uncertainty using the fuzzy set theoretic cloud model.

An example of collecting information while reading papers.

Literature Review

Identifying the challenge and establishing the need to develop methods to deal with it.

Have a purpose, drive the reader towards your contributions, critically assess past contributions to highlight why these aspects are important.

Simple illustration of why an existing metric used in the field is not appropriate.

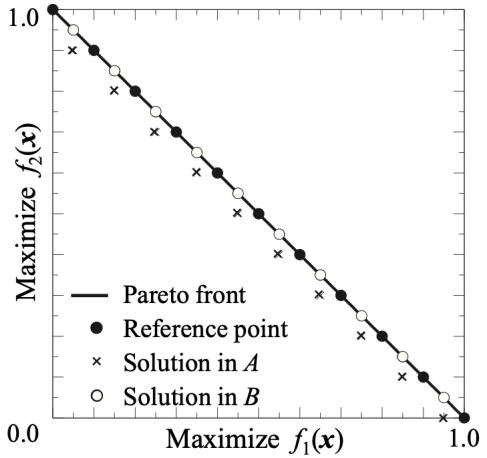
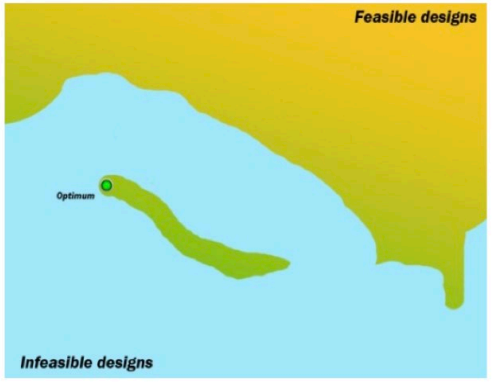


Fig. 3. Reference points ($H = 10$) and two solution sets A and B.

Could be a schematic/ conceptual diagram too.



Disconnected Feasible Regions

Present the problem that is easy to comprehend. Illustrate what happens when existing methods are used to solve the problem. Indicate what is the theoretical solution and explain why existing methods fail to deliver solutions that are close to theoretical solutions.

What needs to be done to solve such problems and present your pitch.

Summary

Use keywords and search across databases e.g. SCOPUS, Google Scholar or specific journals.

There may already be review papers which are good starting points. Does your review include latest relavent papers ?

Often students pick up random articles which they think is valuable. It is increasingly becoming a problem and hence first point of check is "Does the paper appear in top 3 journals in the field ?"

Store the PDF and the bibliography as you go. May use JabRef. Bibliography picked up online might not be complete, so please make sure and fill up fields if they are missing.

Literature review has a purpose. Reading papers is good but if you are not doing it efficiently, you would be wasting a lot of time. Create an excel spreadsheet and populate information as you go with an aim to classify things. Often papers talk about strengths and its your job to **dig deep to find out their limitations.**

Having a review **without a structure** and talking about papers one after another in a random order is useless. Similarly, citing without discussion is useless.

Summary

Listing a set of important research questions is one of the greatest challenges in research, especially if you are new to the area.

Too broad a question. Too narrow or restricted focus. Hypothesis cannot be tested. Always think about what will your method deliver that one cannot get today. Why does anyone need your method ? Is your method too restricve in scope ? Are elements novel ? How do you plan to demonstrate the performance ?

Claims: Be careful to limit your claims.

Codes and strategies will evolve with time. Just because you have written a code and generated results does not mean it is cast in stone. You will need to rewrite codes or strategies over and over again, so following a good coding practice is necessary. Often scripting is the way to go that automatically runs problems and generates results/tables/figures and the whole lot. Yes, it is more time consuming the first time, but saves heaps of time in long run.

Writing pseudocodes and generating flowcharts and illustrations are extremely important. Here is an example https://science.sciencemag.org/content/324/5923/81.abstract

Avoid the temptation to coin your own terms when something already exists to describe that.

Additional Information

What is top quality ? Usually top 5 in Q1 list.	https://www.scimagojr.com
Creating a comprehensive and complete bibliography is useful.	https://www.jabref.org
Managing and editing drafts.	https://www.overleaf.com
What is a good review paper ?	There could be many review papers. Look up on the number of citations, reputation of the author (I expect them to be pioneers in the field), how old is it ? https://www.sciencedirect.com/science/article/abs/pii/S2210650211000198
There are often tutorial slides by pioneers talking about the field.	https://link.springer.com/chapter/10.1007/978-3-642-17144-4_1

Garbage Dilemma

A paper that clearly explains the problem. It's a recent paper too.

Highly complex algorithm with millions of components/strategies.

Millions of parameters.

Results compared on a million other problems and a million other algorithms.

Millions of tables and graphs and statistical analysis of all sorts.

Still does not appear in top 3 Q1 outlet.

Does not provide a rationale as why the algorithm delivered better results when compared with others.

Results presented in Table 1 indicates that the mean/median performance of the algorithm is significantly better than all others across all problems.

There is no explanation as to why your proposed algorithm performed better ?

There is no indication on what classes of problems the proposed algorithm will face challenges with poor performance.

There is no justification on the design of your schemes and rationale on the choice of the parameters.

Scheme switch on/off might be necessary to show and explain the need of the scheme. Again, the focus should be on interpretation and don't just point to results or tables.

Research Presentation: General Guidelines

Please use consistent fonts. Title, Section, Subsection etc etc should be larger and consistent across the presentation.

Number of slides/pages in a paper is like real estate. Use it wisely.

Avoid too much text. Use schematics/Figures to your advantage. Go slow with the first few slides, spend time to get the key message across so that the audience understands your problem.

Slide-1: Big picture first: What is the problem you are trying to solve ? Why is it important should be stated.

Slide-2: Surely, you are not the first to study this problem. Indicate who attempted to solve this problem. Good to indicate dates, e.g. (Ray, 2021), Singh(2019) etc etc that shows you have looked into something that is recent. You should classify such attempts. For example, single objective formulation, multiobjective formulation etc etc.

Slide-3: If there is past research in this area, what are the problems with those approaches ? Talk about the drawbacks of the approaches you listed in Slide-2 to build a case for the need to improve performance or develop new capability to deal with the challenges.

Research Presentation: General Guidelines

Slide-4: What are the 3 or 4 aims of your research

Slide-5: What is novelty and significance of your approach linked to the aims.

Slide-6: What is your methodology to deal with Aim 1

Slide-7: What is your methodology to deal with Aim 2

Slide-8: What is your methodology to deal with Aim 3.

For slides 6, 7 and 8 make sure you indicate how you plan to compare your performance with what and based on what metrics..

Slide 9: What have you achieved. Ideally you had been working on the first aim. Again linking is important to aim.

Slide 10: Timeline of planned activities

Slide 11: References

Statements and Claims

They should be technical and not ambiguous. In this research, an efficient algorithm will be designed to deal with large-scale, real-life instances of XYZ. Such a statement in itself does not make any sense. What is real life ? What is large scale ?

Cut copy and paste equations: It shows you are lazy or don't know how to write equations in an editor. Be mindful of the fonts in the equations too. Vectors and scalars representation nomenclature.

Having figures with illegible legends, axis titles is an indication you have poor presentation skills. Once again consistent size of figures and resolution is important.

Presenting a whole stack of results in tables is useless unless you can draw attention to things in the table right away in 2 secs.

Limitations: Don't be afraid to state the limitations of your approach. Nothing is perfect.

Finally, do you have any recommendations for potential extensions. This reflects that you have the appropriate knowledge to expand the field of research.

Some Styles

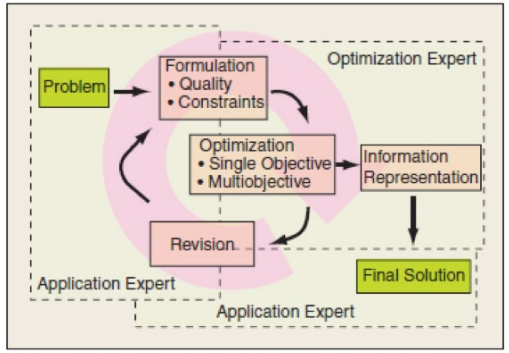


FIGURE 13 A loop for formulation of the optimization problem.

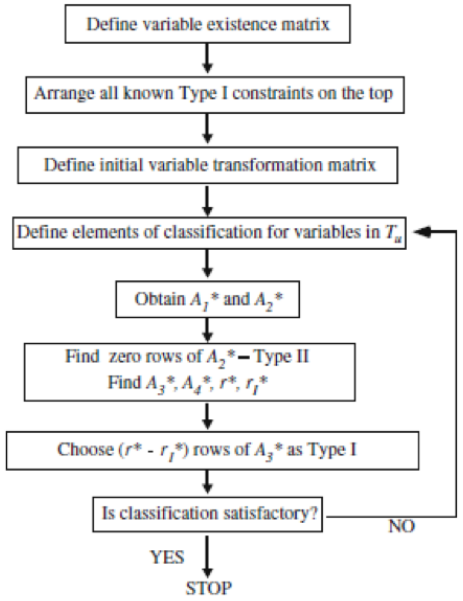


Fig. 3 Proposed interactive classification of equality constraints

Algorithm 4 Niching $(K, \rho_j, \pi, d, Z^r, F_l, P_{t+1})$ procedure

Input: K, ρ_j, $\pi(\mathbf{s} \in S_t)$, $d(\mathbf{s} \in S_t)$, Z^r, F_l
Output: P_{t+1}

$k = 1$
while $k \leq K$ **do**
  $J_{\min} = \{j : \text{argmin}_{j \in Z^r} \rho_j\}$
  $\bar{j} = \text{random}(J_{\min})$
  $I_{\bar{j}} = \{\mathbf{s} : \pi(\mathbf{s}) = \bar{j}, \mathbf{s} \in F_l\}$
  if $I_{\bar{j}} \neq \emptyset$ **then**
    if $\rho_{\bar{j}} = 0$ **then**
      $P_{t+1} = P_{t+1} \cup \left(\mathbf{s} : \text{argmin}_{\mathbf{s} \in I_{\bar{j}}} d(\mathbf{s})\right)$
    else
      $P_{t+1} = P_{t+1} \cup \text{random}(I_{\bar{j}})$
    end if
    $\rho_{\bar{j}} = \rho_{\bar{j}} + 1$, $F_l = F_l \backslash \mathbf{s}$
    $k = k + 1$
  else
    $Z^r = Z^r / \{\bar{j}\}$
  end if
end while

Algorithm 2. MOEA/D using Pareto adaptive scalarising functions

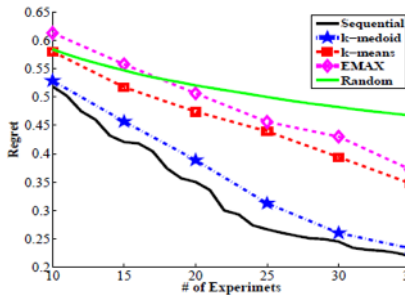
Input: initial population, $S \leftarrow \{\mathbf{x}^1, \mathbf{x}^2, \cdots, \mathbf{x}^N\}$, initial weights, $W \leftarrow \{\mathbf{w}^1, \mathbf{w}^2, \cdots, \mathbf{w}^N\}$, selection neighbourhood size, T, replacement neighbourhood size, nr
Output: S

Initialise the L_k^p as the weighted sum style, i.e., $p_k \leftarrow 1, i \in \{1, 2, \cdots, N\}$;
Evaluate the objective function values of the initial S;
Update the *ideal* and *nadir* vectors, $\mathbf{z}^*$ and $\mathbf{z}^{nad}$;
Randomly assign each weight, $\mathbf{w}^i$ with a candidate solution, $\mathbf{x}^i$;
Calculate the Euclidean distance between weights, $\mathbf{w}^i$ and $\mathbf{w}^j$, $i, j \in 1, 2, \cdots, N$;
Find the T neighbouring weights $B(\mathbf{w}^i)$ of $\mathbf{w}^i$ based on the distance of weights and identify the related neighbouring solutions Q of $\mathbf{x}^i$;
Set *iteration* $\leftarrow 0$, set *matingS* $\leftarrow \emptyset$;
while *the stopping criterion is not satisfied* **do**
  for $i \leftarrow 1$ **to** N **do**
    if *rand* $< \delta$ **then**
      matingS $\leftarrow Q$;
    else
      matingS $\leftarrow S$;
    end
    Randomly select three solutions $\mathbf{x}^{r1}, \mathbf{x}^{r2}$ and $\mathbf{x}^{r3}$ from the mating pool, *matingS*;
    Generate a new solution $\mathbf{x}^{new}$ by performing differential evolution (DE) and polynomial mutation (PM) operators;
    Evaluate the objective value of $\mathbf{x}^{new}$, and update the ideal and nadir vectors;
    for *each* $\mathbf{x}^k \in Q$ **do**
      Compare $g^{wd}(\mathbf{x}^{new}|\mathbf{w}^k, p^k)$ with $g^{wd}(\mathbf{x}^k|\mathbf{w}^k, p_k)$;
    end
    Replace no more than nr solutions in Q with $\mathbf{x}^{new}$ if $g^{wd}(\mathbf{x}^{new}|\mathbf{w}^k, p_k)$ is smaller;
  end
  Update the p_k value for each search direction using Algorithm 1;
end

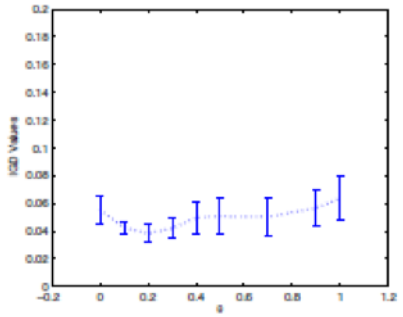
Both of them are valid. It depends on the outlet and the domain.

Some Styles : Watch Closely

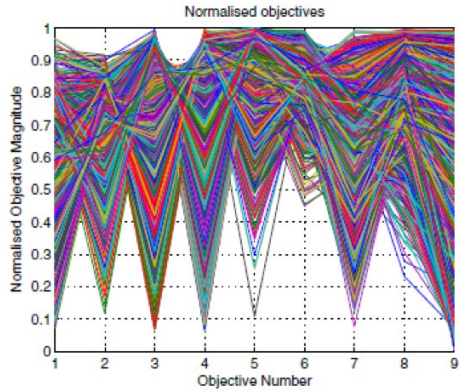
What is the best way to present the results ?



Fuel Cell



(a) DTLZ1



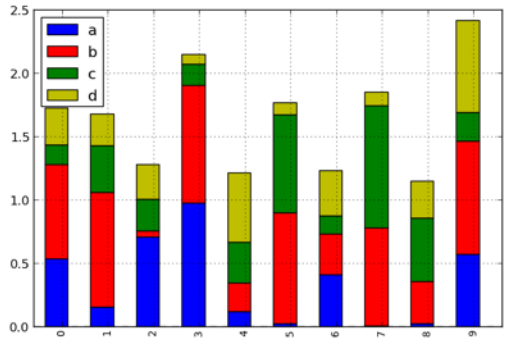
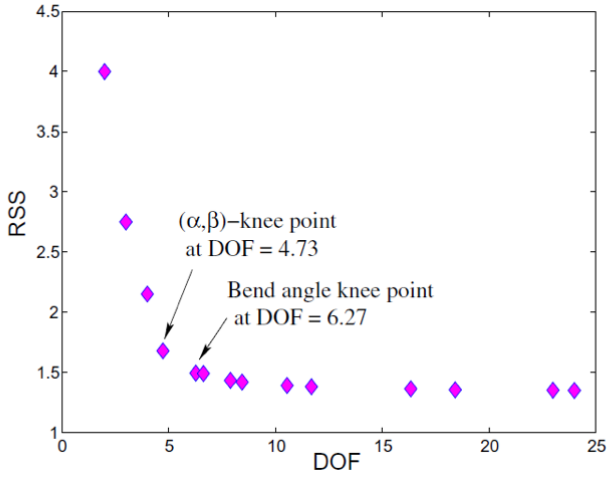
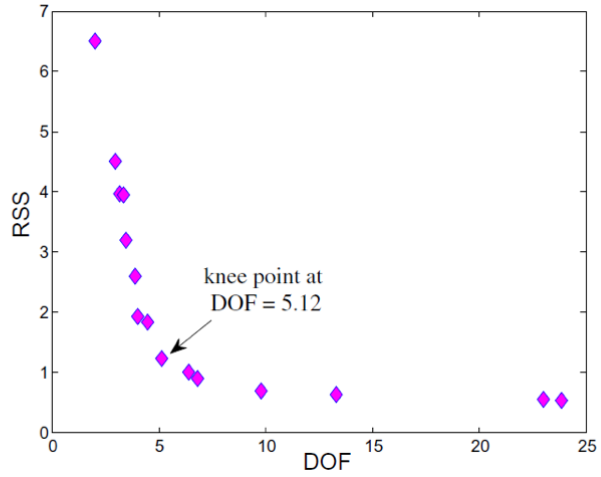


Fig. 6. Plot of 5000 non-dominated objective vectors. Objectives are normalised so that full 775,140 point vector set lies in range [0,1].





HV VALUES BASED ON 50 INDEPENDENT RUNS

Algorithm	FE	Hypervolume			
		Best	Mean	Worst	Std
I-DBEA	50,000	**0.02995**	**0.01726**	**0.00699**	0.05121
DBEA-Eps		0.02899	0.01715	0.00689	0.04561
ϵ-MOEA		0.02032	0.01032	0.00259	0.04125
Borg-MOEA		0.02245	0.01013	0.00424	0.02327
MOEA/D-DE		0.00092	0.00087	0.00045	0.00145
ϵ-NSGA-II		0.01636	0.01005	0.00236	**0.05232**

Think about using markers, line styles, color appropriately. Think about what would happen if the draft is printed in BW.