Surveying Design Spaces with Performance Maps

A Multivariate Visualization Method for Parametric Design and Architectural Design Optimization

Thomas Wortmann
Singapore University of Technology and Design
thomas_wortmann@mymail.sutd.edu.sg

This paper presents a novel method to visualize high dimensional parametric design spaces with applications in computational design space exploration. Specifically, the visualization method presented here supports the understanding of design problems in architectural design optimization by allowing designers to move between a high dimensional design space and a low dimensional "performance map". This performance map displays the characteristics of the fitness landscape, develops designers' intuitions about the relationships between design parameters and performance, allows designers to examine promising design variants and delineates promising areas for further design exploration.

Keywords: Fitness Landscape, Design Space Exploration, Multivariate Visualization, Optimization, Star Coordinates

A NEED FOR DESIGN SPACE CARTOGRAPHY

The challenge of representing high-dimensional data in the two or three dimensions that are visually understandable by humans appears in many disciplines, and - with the advent of Big Data in the last decade - has emerged as a field of study in its own right. This paper presents a novel method to visualize high dimensional parametric design spaces and their performance - such as efficiency in energy and material consumption - with applications in computational design space exploration. Specifically, the visualization method presented here supports the understanding of design problems in architectural design optimization (ADO) by allowing designers to move back and forth between a high dimensional design space and a low dimensional "Performance Map". Architectural design space exploration and architectural design optimization rely on the notion of design spaces to organize design variants and processes (Woodbury and Burrow 2006). Since these spaces often have more than two or three dimensions, they are difficult to visualize. In parametric design, the number of dimensions of the design space equals the number of design parameters. The difficulty of representing such abstract spaces motives the need for multivariate visualizations, which, for example, helps designers understand the relationships between parameters and the similarities characterizing groupings of design variants.

ADO combines the notion of a parametrically defined design space with one or more numerically expressed performance criteria. In ADO, performance criteria typically pertain to reducing a building de-
sign's resource or energy consumption. For example, the dome of the Louvre Abu Dhabi - with a structural span of 165 meters - has been optimized both in terms of its structural and daylighting performance (Imbert et al. 2013). When such performance criteria are included in a design space, the literature often speaks of a fitness landscape. This paper focuses on representing fitness landscapes with only one performance criterion. (A possible extension to multiple criteria is discussed in the conclusion.) In a fitness landscape, the height of a coordinate in the fitness landscape corresponds to the performance value of a design variant, and the remaining coordinates specify the design itself. As an extension of a design space, such a fitness landscape is often defined by more than two parameters, which makes it impossible to represent as a literal landscape.

Recently, ADO has received new understandings both as a generative design tool that provides starting points for further design exploration (Bradner et al. 2014) and as a representational tool that aids the understanding of design problems (Wortmann et al. 2015). Chen et al. (2015) attempt to group large numbers of design variants - found with a genetic algorithm - with a clustering method to better understand the relationship between design features and performance. Their effort is symptomatic of the need for human-understandable representations of fitness landscapes. In other words, ADO requires simple (low dimensional) representations of complex (high dimensional) relationships. This paper presents a novel, low-dimensional representation of a high dimensional fitness landscape - the Performance Map that has at least four applications: It provides insights into the characteristics of the fitness landscape, and thus into the nature of the optimization problem (1), develops designers' intuitions about the relationships between design parameters and performance (2), allows designers to examine promising design variants (3) and delineates promising areas for further - manual or automatic - design exploration (4). This paper shows how to construct such a performance map, followed by an example and comparison with an existing technique and a more extensive discussion of potential applications. The next section discusses multivariate visualizations relevant to design space exploration and ADO and addresses the reversibility of these methods, a property that is necessary for turning them into exploratory, interactive design tools.

**MULTIVARIATE VISUALIZATIONS**

Multivariate visualizations represent data that have more than two or three dimensions and thus are difficult to display on print-outs and screens. They serve a range of purposes, for example summarizing data, identifying patterns in or similarities between data, and displaying correlations between parameters. Hoffman and Grinstein (2002) survey multivariate visualization types. According to them, multivariate visualizations harness humans' perceptual abilities to "look for structure, features, patterns, trends, anomalies, and relationships in data" (p. 21). De Oliveira and Levkowitz (2003) describe how "[multivariate] visual mapping techniques are now being used both to convey results of [data] mining algorithms in a manner more understandable to end users and to help them understand how an algorithm works." They suggest that multivariate visualization makes data mining methods more transparent and interactive. This paper proposes that design space exploration and ADO can benefit from multivariate visualization in a similar manner, by, instead of presenting only one or a small selection of design variants, providing the designer with an overview of the design space or fitness landscape.

**Parallel and Radial Visualizations**

A straightforward method to represent data with several parameters is Parallel Coordinates. Parallel Coordinates introduces a set of parallel - usually vertical - axes equal to the number of parameters for the data. To display a datum, one marks the value of each parameter on the corresponding axis and connects the resulting points with a polyline (see Figure 1). A variation of this method uses radial axes, which allow the
Parallel Coordinates represents each of the 362 explored designs of the example optimization problem as a polyline, with the best-performing design variants on top.

The visualization method applied in this paper - Star Coordinates - also introduces one coordinate axis for each parameter, with the axes typically arranged radially (see Figure 2). In contrast to Parallel Coordinates, however, Star Coordinates displays a datum not as a closed polyline but as a single point. As a point-based representation, Star Coordinates can represent many more data than Parallel Coordinates without overlaps, and even display a continuous field to represent the space of the data, although at the price of making individual parameters less readable. Note that, although in theory Star Coordinates might represent several data with an identical point, the Performance Map circumvents this limitation by preferring better performing design variants.

RadViz - a method closely related to Star Coordinates - places data with a spring system, with each parameter represented by a spring. However, Rubio-Sanchez et al. (2016) conclude in a comparison of the two methods that RadViz "introduces non-linear distortions, can encumber outlier detection, prevents associating the plots with useful linear mappings, and impedes estimating original data attributes accurately."

An important advantage of parallel and radial visualizations is that their coordinate axes allow the estimation of numerical parameters. This advantage contrasts with other multivariate visualization methods such as Self-Organizing Maps (Kohonen 1982), which cannot directly represent numerical relationships. The ordering of coordinate axes can influence the insightfulness of parallel and radial representations. The figures presented here order the coordinate axes according to the order of the parameters in the underlying optimization problem.

Visualizing Optimization Results
In the optimization field, fitness landscapes are usually represented with a Matrix of Contour Plots. Such a matrix consists of $n^2 - n$ two-dimensional plots, with each plot representing the relationship between a pair of parameters. The plots on the diagonal of the matrix are left blank or show the relationship of the performance objective to a single variable (see Figure 4). In other words, this visualization consists of two-dimensional "sections" or "contours" through the higher dimensional space. This architecturally-inspired metaphor clarifies an important limitation: Although the number of contour
plots increases quadratically with the number of parameters, these plots display only a very small portion of the space, since, to draw a single plot for two parameters, all other parameters must be kept constant. In terms of the section metaphor, the sections are drawn through a single "base point" in the higher dimensional space. This limitation is usually mitigated by basing the visualization on the best set of parameters found during optimization. Van Wijk and van Liere (1993) propose an interactive Matrix of Contour Plots that allows the user to navigate the space by changing the base point of the visualization.

Nevertheless, contour plots are unable to provide an overview of the design space as a whole and, due to their exponentially growing number, become increasingly difficult to understand even for relatively small numbers of parameters. Additionally, there is no guarantee that the selection of design variants displayed is relevant for understanding the design space at hand, since the contour plots only capture interactions between pairs of parameters and thus miss other, potentially more important interactions. This inability to represent non-pairwise, non-linear interactions adds to the cognitive difficulty of integrating information from an exponentially large number of contour plots.

Pareto Fronts are a type of visualization often employed in multi-objective optimization. Such fronts are two- or three dimensional plots of the relationship between two or three performance criteria. Although Pareto Fronts support designers in understanding tradeoffs between different criteria (Radford and Gero 1980), they are of limited use for design space exploration, since they cannot visualize the relationship between performance criteria and design parameters.

Reversing Multivariate Visualizations
An important issue for design space exploration is the reversibility of multivariate visualizations. In order to support designers with interactive tools that visualize design spaces and fitness landscapes, one should not only be able to map high dimensional designs into a low dimensional "performance map" for visualization, but also to map locations on the "performance map" back into the high dimensional design space. Ideally, such a mapping would bijective. In other words, every design in the higher dimensional design space would map uniquely onto a lower dimensional representation, and every lower dimensional representation would map onto exactly one design in the higher dimensional space. Many multivariate visualization methods don't support such a "reverse" mapping from low to high dimensions, and none support a bijective mapping. Matrices of Contour plots allow reverse mappings, but since this method only takes (relatively arbitrary) slices of the higher dimensional space, contour plots are not bijective. Instead, they are injective, with every location in the two-dimensional representation mapping onto a unique design variant, but with many design variants not representable at all.

This paper proposes a novel extension to Star Coordinates that allows designer to go back and forth between the design space and its representation—the Performance Map—using a surjective mapping. A surjective mapping, while falling short of the bijective ideal, is an improvement over injective mappings in that all designs from the higher dimensional space can potentially be represented, but with some design variants overlapping in the representation. The next section details this surjective mapping.

METHOD
The method presented in this section extends the metaphor of the fitness landscape into higher dimensions through a multivariate visualization that represents performance not as elevation but as color. A novel extension of Star Coordinates, the method allows mappings from the two-dimensional Performance Map to the high-dimensional design space. The method achieves this reversible, surjective mapping in four steps: Project the evaluated design variants (i.e. design variants whose performance has been simulated) onto two dimensions using Star Coordinates (1), compute a Delauney triangulation for
the projected points in the two-dimensional map (2) and approximate the performance of unexplored design variants (i.e. design variants with unknown performance values) by interpolating their performance values based on the corner points of the triangulation's triangles (3) or by estimating their performance values via a surrogate model (4). Note that, since the Performance Map is surjective and based on interpolating between evaluated designs, it does not represent the whole design space, but rather a collection of design variants that is relatively similar to already evaluated designs. This similarity, however, increases the likelihood of performance estimates to be accurate. In other words, it is of little use to represent portions of the design space that have not been previously explored at all.

**Star Coordinates**
The method projects evaluated design variants using Star Coordinates. Although other arrangements are possible, the coordinate axes are typically spaced equally around a circle, which is also true for the example presented here. Note that, to avoid skewing the visualization due to differences in scale between parameter values, it is advisable to normalize the parameters to the same range, i.e. [0, 1]. Star Coordinates determines the two-dimensional position $p$ of a $n$-dimensional design by multiplying the value of each parameter with the vector representing the corresponding coordinate axis, and by then adding the resulting scaled vectors. In other words, the two-dimensional embedding $p \in \mathbb{R}^2$ of the $n$-dimensional design $x_n \in \mathbb{R}^n$ is a linear combination of $n$ coordinate vectors $v$:

$$p = x_1 v_1 + x_2 v_2 + \ldots + x_n v_n$$

A potential complication for Star Coordinates is that two designs with different parameters can map onto an identical point. For example, all designs which have the same value for all parameters (all zero, all one, etc.), map onto the origin of the coordinate axes. Since the two-dimensional mapping consists of linear combinations of more than two, pairwise linearly independent vectors in the plane, a two-dimensional point maps to a theoretically infinite number of parameter sets in the higher dimensional space. This mapping to more than one parameter set implies that Star Coordinates is not bijective, but surjective: Every parameter set maps onto a unique point, but this uniqueness is not true in reverse. In terms of creating the Performance Map, this complication is resolved by preferring the better performing design variant when several variants map to the same point. (In practice variants rarely overlap). As a result, the Performance Map is slightly biased towards better performing designs.
Triangulating a Surjective Mapping

The Delauney algorithm (de Berg et al. 2008) provides a quick \(O(m \log m)\) and unique triangulation for \(m\) points in the plane. The triangulated points are termed vertices, the connections between the vertices edges, and the triangles formed by the edges faces. Since, in the Delauney triangulation, every set of 2D coordinates is located in exactly one face (with coordinates coinciding with edges or vertices a special case), the triangulation provides a unique set of three vertices (i.e. three mappings of evaluated design variants) for every set of 2D coordinates in terms of the face’s vertices (see Figure 2). Delauney triangulations have the attractive property that every vertex is connected to its nearest neighbor by an edge. As a consequence, a point in the Performance Map will always be close -though not necessarily closest- to the three vertices of its face, i.e. it will always relate closely to the three evaluated design variants that its parameters are interpolated between. An alternative is to find the three nearest neighbours for every point, which, however, is more computationally expensive and less visually intuitive.

Interpolating Performance Values

The method approximates the performance of unexplored design variants by linearly interpolating between either the performance values or the design parameters of the three evaluated designs associated with an unexplored variant through the Delauney triangulation. In the latter case, the method obtains an estimate from a surrogate model using interpolated design parameters. The linear interpolation represents each evaluated design as a three-dimensional point \(v\), where the first two coordinates are given by Star Coordinates and the third coordinate is the performance value. The approximated performance \(p_z\) of the unexplored variant is given by the intersection of a vertical line through the 2D coordinates of the variant \(p\) and the plane of the three three-dimensional points \(v\) of the evaluated designs forming the corresponding face of the triangulation (2). The plane is defined by calculating its normal \(n\) (1):

\[
\begin{align*}
    n &= (v_1 - v_0) \times (v_2 - v_0) \\
    p_z &= \frac{-n_x \cdot p_x - n_y \cdot p_y + n \cdot v_1}{n_z}
\end{align*}
\]

Estimating Performance Values

A more sophisticated method to approximate the performance of unexplored design variants is the employment of a surrogate model. Surrogate models approximate unexplored fitness landscapes from evaluated designs by using machine learning techniques. They interactively approximate performance values and speed up optimization processes by either supplementing or completely supplanting time-intensive simulations of design variants (Koziel and Leifsson 2013). The surjective mapping presented here employs the linear interpolation described above to interpolate not between the performance values, but between the design parameters of evaluated designs. Using the interpolated parameters, a surrogate model quickly generates a performance estimate that is more accurate than the linear interpolation described above. In the example presented in the next section, a model-based optimization algorithm has iteratively (re-)constructed the surrogate model during the optimization process to decide which design variant to evaluate next (Wortmann et al. 2015).

Visualizing an Example Problem

This section presents a Performance Map created with the method described above. The Performance Map visualizes a standard problem from structural engineering that concerns a small transmission tower to be optimized in terms of weight while meeting stress and deflection constraints. The tower consists of 25 structural members, which are categorized into eight beam groups. One of thirty cross sections can be assigned to each beam group. The problem thus has eight variables that can take one of thirty integer values. Consequently, there are \(30^8\) design variants. The visualizations presented here represent...
The Performance Map represents the design space in terms of the estimated performance values of unexplored design variants and indicates explored variants. Note the groupings of well-performing variants in the upper left corner and near the origin.

The optimization results and surrogate model derived from a ten-minute run with RBFOpt, a model-based optimization algorithm. During ten minutes, 362 designs were evaluated through a structural simulation, with the lightest design weighing 240 kilograms. (For details on the benchmark problem and optimization method see (Wortmann and Nannicini 2016).)

**Performance Map**

In Figures 2 and 3, every evaluated design is represented as a colored circle, with the position of each circle corresponding to the design's parameters, and the circle's fill color corresponding to its weight. Due to the large weight difference between the best and the worst evaluated designs (240 kg and 1940 kg) -which also includes penalties for exceeding stress and deflection constraints- the colors are applied on a logarithmic scale. In this way, more detail is visible for the better, lighter designs. The colors follow a palette proposed by Niccoli and Lynch (2012) that is more perceptually adequate than conventional rainbow palettes due to its linearly increasing luminance. The color palette and scale are identical for all figures.

Figure 2 indicates only the colored points of the evaluated designs on the left, with the Delauney triangulation used to interpolated between them on the right. Figure 3 shows the approximated Performance Map derived from the surrogate model. Note how in Figure 3, the visualization indicates two sepa-
rate areas of good, i.e. light designs: One closer to the center, and one in the upper left corner. These groupings indicate two different types of solutions, with each type displaying similar parameters. The visualization also shows that the central group has been explored extensively, while the peripheral group consists only of a small number of explored points. The visualized surrogate model indicates that this area might contain very high-performing design variants and thus is promising for future exploration.

**Comparison with Matrix of Contour Plots**

The contour plots in Figure 4 represent the identical optimization results and surrogate model as Figure 3. Note that, in this visualization, only the best optimization result can be indicated because the contour plots do not contain the remaining evaluated points. Also note that the two different areas of good designs visible in Figure 3 are not identifiable in Figure 4. This comparison between Star Coordinates and a Matrix of Contour Plots confirms the disadvantages of contour plots identified above, such as the fragmentation of the design space into a quadratic number of contours and the inability to identify nonlinearities between more than two parameters. The next section discusses potential applications for the Performance Map, with an emphasis on interactive design space exploration and optimization.
APPLICATIONS
The proposed visualization method not only supports the designer’s understanding of the fitness landscape, but also promises to improve the interactivity of ADO tools. Performance maps provide insights into the nature of the optimization problem (1), represent the predicted performance of unexplored designs (2), display designs that are estimated to perform well (3), and identify promising areas for further exploration (4). Beyond ADO, applications extend to higher dimensional design space exploration more broadly. For example, the mapping can represent similarities between parametrically defined designs or display trajectories of design decisions in a parametric design space.

Identifying the Optimization Problem
Performance Maps capture salient characteristics of the fitness landscape. For instance, a fitness landscape is "smooth" when the relationships between design parameters and the performance criterion are relatively linear and without discontinues, and "rugged" when they are not. Another important distinction is between "convex" fitness landscapes with only a single peak (i.e. optimum) and "non-convex" landscapes with additional peaks (i.e. sub-optima). Such characteristics, which define the difficulty of an optimization problem and inform the selection of an appropriate optimization algorithm, can be identified via a Performance Map. For example, the performance map in Figure 3 shows multiple sub-optima and a rugged fitness landscape where design variants with similar parameters can have very different performance values. This ruggedness is expected for a structural optimization problem, where the lightest, feasible designs are similar to failing ones.

Relating Parameters and Performance
Both the second and third possibilities imply linking the performance map to the underlying parametric model in real-time. Designers can adjust the parametric model according to design intentions not captured by a numerical performance value and simultaneously obtain the estimated performance of the adjusted design as a location on the Performance Map. They can also identify similarities between the new design variant and evaluated ones in terms of their locations in the two-dimensional parameter space of the map. If the new design variant is promising, they can evaluate its performance and add this information to the Performance Map and underlying surrogate model.

Examining Promising Designs
Conversely, the designer can select a promising location on the Performance Map and simultaneously see its representation in the original parametric model. In this way, the designer can appreciate the values for the design parameters that correspond to a location in the performance map, and what kind of design these parameters imply. A simple but important part of fostering this appreciation is to facilitate zooming and panning over the Performance Map.

Guiding Automated Design Exploration
Finally, based on the understanding gained from the Performance Map, the designer can guide further design exploration by limiting the range of the design parameters. This limiting can be achieved by selecting a promising area on the Performance Map and restricting future exploration to the parameter ranges in the selected area. By selecting a specific parameter range, a designer can point the optimization algorithm in a direction that supports her design intentions while ensuring a more efficient optimization process due to the smaller design space to be explored. In short, Performance Maps allow designers to interact manually with an otherwise automated optimization process.

Conclusion and Future Research
Beyond presenting a larger number of examples and implementing the possibilities described in the previous section, several directions present themselves for future research. Assuming that a surrogate model of the performance dimension is available, one can statistically analyse the model’s performance values to identify correlations between parameters and the
sensitivity of the parameters in terms of performance. Such analyses can increase the relevance of visualizations by suggesting orderings for the coordinate axes or a reduced selection of parameters to be displayed.

Another direction is the visualization of multiple performance criteria. Heinrich and Ayres (2016) have overlaid different color palettes to visualize multiple performance criteria using contour plots. Their approach implies creating a performance map for each criterion individually, which from a mathematical optimization perspective is much easier than trying to address multiple criteria at once. Overlaying the results from several single criteria in a visual representation therefore not only promises greater understanding, but also better designs found in a shorter amount of time. Combining their approach of overlaying color palettes with the Performance Map presented here should yield insightful visualizations that represent the design space in its relationship to multiple performance criteria.

The Performance Map presented here fits into a wider context of framing ADO not as a methodology to find the "best" performing solution, but to better understand the relationships between design parameters and one or more numerical performance criteria. It demonstrates the importance of multivariate visualization for fostering better understandings of design spaces, both with and without numerical performance criteria.

Acknowledgment
I thank Prof. Nannicini (SUTD) for his constant advice, and especially the inclusion of features in RBFOpt (projects.coin-or.org/RBFOpt/wiki) that facilitated this research.

REFERENCES

Bradner, E, Iorio, F and Davis, M 2014 'Parameters tell the design story: ideation and abstraction in design optimization', Proceedings of SIMAUD
Heinrich, MK and Ayres, P 2016 'For Time-Continuous Optimisation', Living Systems and Micro-Utopias: Towards Continuous Designing, Proceedings of the 21st CAADRIA, Hong Kong, pp. 373-382
Koziel, S and Leifsson, L (eds) 2013, Surrogate-Based Modeling and Optimization, Springer, New York, NY
Niccoli, M and Lynch, S 2012 'A more perceptual color palette for structure maps', CSEG GeoConvention 2012Vision
van Wijk, JJ and van Liere, R 1993 'HyperSlice: Visualization of Scalar Functions of Many Variables', Proceedings of the 4th Conference on Visualization ’93, Washington, DC, USA, pp. 119-125
Woodbury, RF and Burrow, AL 2006, 'Whither design space?', AIEDAM, 20(2), pp. 63-82
Wortmann, T and Nannicini, G 2016 'Black-box opt. for architectural design', Living Systems and Micro-Utopias: Towards Continuous Designing, Proceedings of the 21st CAADRIA, Hong Kong, pp. 177-186