Visualising high-dimensional Pareto relationships in two-dimensional scatterplots

EMO 2013

Jonathan Fieldsend & Richard Everson

Computer Science, University of Exeter, UK

The problem

- We visually *absorb* data represented in two/three dimensions.
- It is "easy" to detect Pareto domination relationships between points in a set (a population) with two/three objectives plotted as a scatter plot
- Many EMO problems we confront with have more than two/three objectives
- If we would like to represent the population as points on the plane, can we still convey Pareto relationships between members usefully, and detect structures that exist in the original data?

The problem

- What we are concerned with here:
  - Can we develop a scatter plot representation which *loses* as little Pareto relationship information as possible when converting from D-dimensions to 2-dimensions
  - In doing this, can we infer useful/interesting relationships in the higher dimensional space that may be lost/obscured with other mappings?
- What we **are not** concerned with here:
  - Data which is a non-dominated set – we are not interested (here) in visualising a set in which *all* elements are non-dominated with respect to one another.

Loss of information in dimension reduction

- To project an objective vector $\mathbf{y} \in \mathbb{R}^D$ into $\mathbb{R}^2$ we must utilise a dimension reduction technique of some form, and, unless there are redundant or perfectly correlated objectives, some information loss is inevitable.
- There are a number of *general* dimension reduction techniques:
  - PCA
  - ICA (essentially assuming it is over determined)
  - Neuroscale
  - MDS
  - Isomap
  - Radviz
  - ...
- However this are ignorant of (or do not attempt to convey) Pareto relationships between points

Desired properties

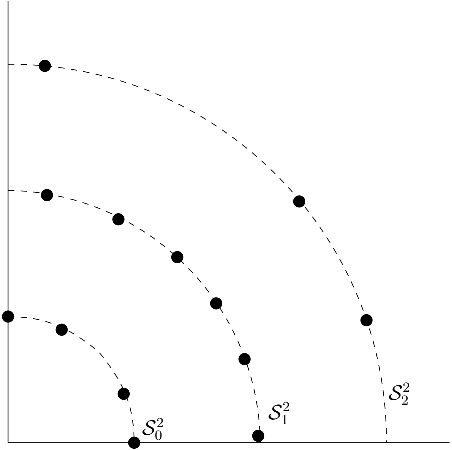
- Given a set $Y^D = \{\mathbf{y}_i\}_{i=1}^N \subset \mathbb{R}^D$
- Want to find a mapping to $Y^2 = \{\mathbf{u}_i\}_{i=1}^N \subset \mathbb{R}^2$ such that if $\mathbf{y}_i \prec \mathbf{y}_j$, then $\mathbf{u}_i \prec \mathbf{u}_j$
- and if $\mathbf{y}_i \not\prec \mathbf{y}_j$, then $\mathbf{u}_i \not\prec \mathbf{u}_j$
- In general a mapping $\mathbf{u} = \mathbf{g}(\mathbf{y})$ with this property does not exist (see e.g. proof in Köppen and Yoshida (2007))

Realistic properties

1. Ensure that the mapping preserves Pareto shells. That is, if we denote by $\mathcal{S}_i^D$ the ith Pareto shell in an ambient space of D dimensions, then $\mathbf{u} \in \mathcal{S}_i^2$ (where $\mathbf{u} = \mathbf{g}(\mathbf{y})$).
2. Minimise dominance misinformation. (We'll look at how to quantify this in later slides.)

(The superscript on $\mathcal{S}_i$ denotes the dimensionality of the space which it inhabits.)

Property 1 is simple to ensure



Property 2 is less simple

- Property 2 requires a direct geometric transference of the dominance relation
- A mapping which attempts a version of this second property is that described in Köppen and Yoshida (2007), which we will briefly describe and illustrate before discussing one of two new approaches in the paper that attempt to tackle both properties

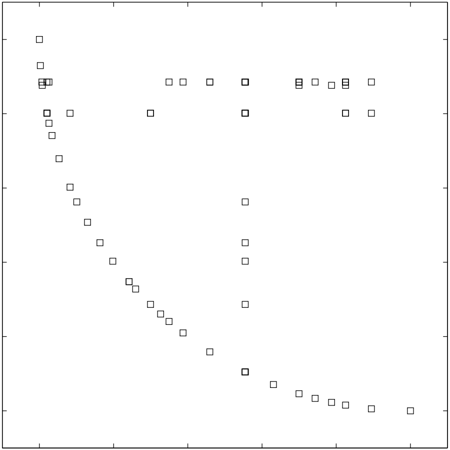
Visualisation of Köppen and Yoshida

- The $\mathbf{S}_0^D$ elements are extracted first and mapped to $\mathbf{S}_0^2$ such that separation between points is proportional to the distances of immediate $\mathcal{S}_0^2$ neighbours in the original $\mathbb{R}^D$ space.
- The mapping has two objectives
  1. The total distance between ordered members in $\mathbb{R}^2$, when projected back into $\mathbb{R}^D$ and traversed should be minimised.
  2. Have a minimal number of instances where an element is placed anywhere *between* two other elements in the ordering, where it does not dominate a set member that the other two *both* do.
- This permutation is optimised using NSGA-II

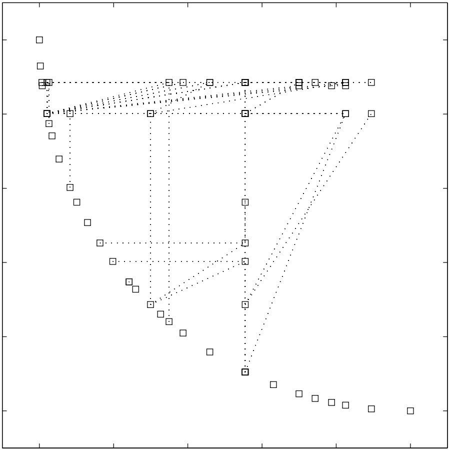
Visualisation of Köppen and Yoshida

- Once the permutation has been optimised, the elements in the non-dominated subset of Y^D can be mapped to $\mathbb{R}^2$
- After this, for every *dominated* point $\mathbf{y}_i \in Y^D$, the subset of $\mathcal{S}_0^D$ which dominates it is determined, and the worst objective values in the mapping of this set are used to fix the position of $\mathbf{u}_i$ in two dimensions.
- Illustration on 100 4-dimensional points sampled at random (from a multivariate Normal distribution)

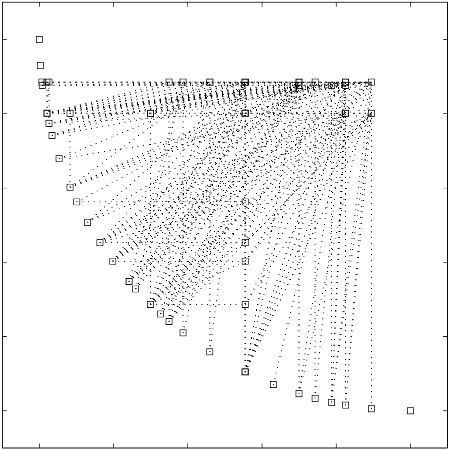
Visualisation of Köppen and Yoshida



Visualisation of Köppen and Yoshida



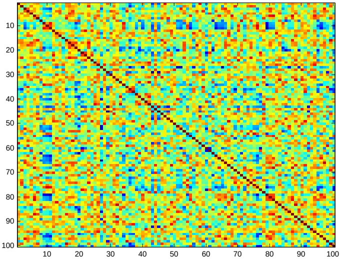
Visualisation of Köppen and Yoshida

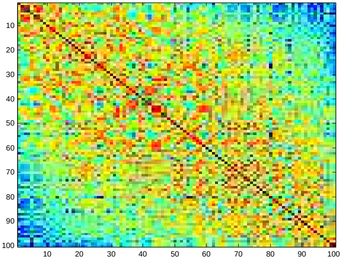


New visualisations

- Shell membership is maintained and clear
  - Distance-based: the *closer* you are to a point, the more likely you are to be dominated
  - Dominance-based: if a point is dominated in one space, it dominates in the mapped space
- Still requires initial mapping $\mathcal{S}_0^D$ to $\mathcal{S}_0^2$
  - Instead of casting this as a problem to tackle with an evolutionary optimiser, we instead order the solutions using spectral seriation (where an approximation can be solved directly using linear algebra).
  - We use *dominance similarity* in the seriation.
  - The dominance similarity between two solutions $\mathbf{y}_j$ and $\mathbf{y}_k$, relative to a third solution $\mathbf{y}_p$, is defined as being proportional to the number of objectives on which $\mathbf{y}_j$ and $\mathbf{y}_k$ have the same relation (greater than, less than, or equal) to $\mathbf{y}_p$.

Seriation of similarity matrices

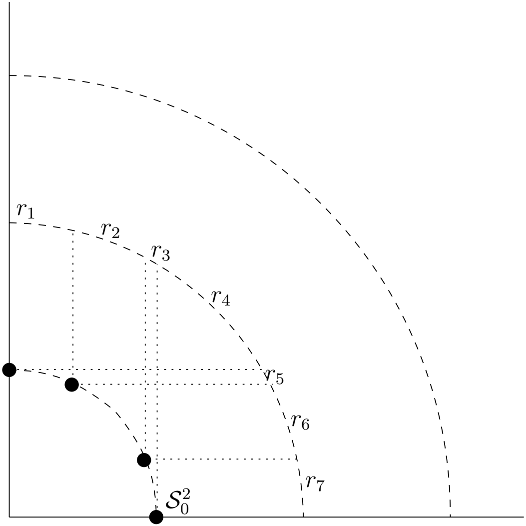




Representing dominance in $\mathbb{R}^D$ by *dominance* in $\mathbb{R}^2$

- if an individual $\mathbf{u} = \mathbf{g}(\mathbf{y})$ has the relationship $\mathbf{y}' \prec \mathbf{y}$, then as far as possible we would like $\mathbf{u}' \prec \mathbf{u}$ to hold (and vice versa).
- We propose a deterministic iterative procedure which attempts to arrange the solutions in each $\mathcal{S}_i^2$ to accomplish this.
- When deciding on the placement of the $\mathcal{S}_i^2$ individuals, the members of $\mathcal{S}_{i-1}^2$ effectively delimit a number of regions on the feasible curve for $\mathcal{S}_i^2$.
- Any point in one of these regions has an equivalent dominance relation with $\mathcal{S}_{i-1}^2$; that is, any point in a particular curve segment r_k is dominated by the same subset of $\mathcal{S}_{i-1}^2$.

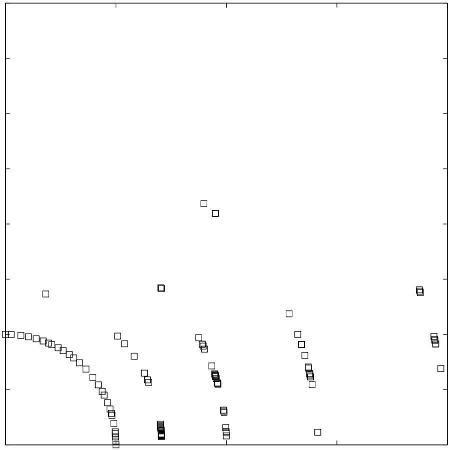
Representing dominance in $\mathbb{R}^D$ by *dominance* in $\mathbb{R}^2$



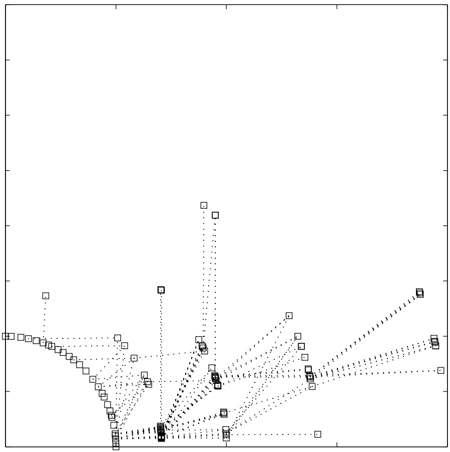
Representing dominance in $\mathbb{R}^D$ by *dominance* in $\mathbb{R}^2$

- We find the regions in the *i*th front, where an element in $\mathcal{S}_i^2$ placed on these line segments is dominated by all the members in lower numbered shells which it is *also* dominated by in the D-dimensional space.
- Amongst these subset of regions, assign the element to the one which has the lowest number of *incorrect dominations*.
- When multiple points are assigned to the same region, distribute them equally across the line segment
  - Ordered according to the their order from spectral seriation.

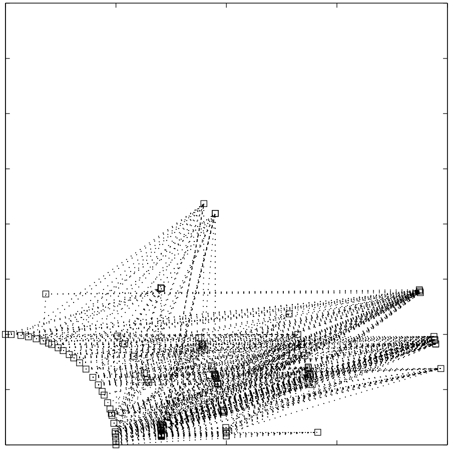
Representing dominance in $\mathbb{R}^D$ by *dominance* in $\mathbb{R}^2$



Representing dominance in $\mathbb{R}^D$ by *dominance* in $\mathbb{R}^2$



Representing dominance in $\mathbb{R}^D$ by *dominance* in $\mathbb{R}^2$

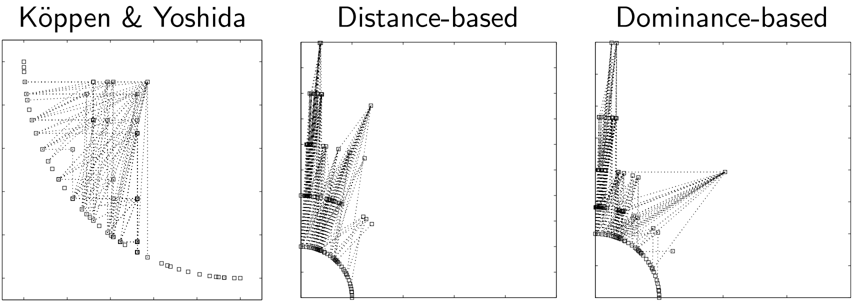


Scatter plot properties

	K & Y	Dist.-based	Dom.-based
(I) If $\mathbf{y}_i \in \mathcal{S}_k^D$ then $\mathbf{u}_i \in \mathcal{S}_k^2$	✗	✓	✓
(II) If $\mathbf{y}_i \prec \mathbf{y}_j$ then $\mathbf{u}_i \prec \mathbf{u}_j$	✗†	✗	✓
(III) If $\mathbf{y}_i \nprec \mathbf{y}_j$ then $\mathbf{u}_i \nprec \mathbf{u}_j$	✗	✗	✗

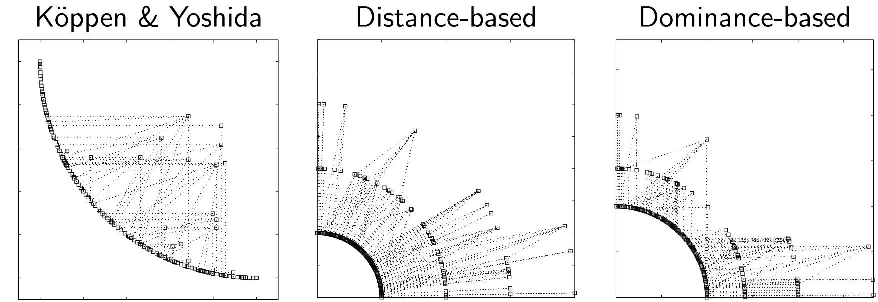
- † If solutions in $\mathcal{S}_0^2$ can be arranged so that second criteria optimised in the approach is equal to zero, then (II) holds for any pair of points which are not mapped to the same location in $\mathbb{R}^2$.

SPEA2 combined archive and search populations, DTLZ2, 4-objectives



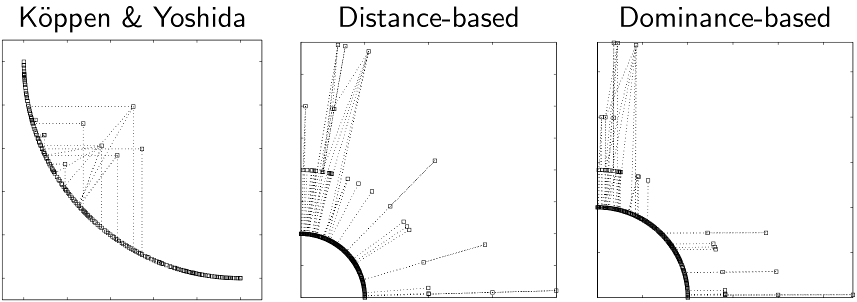
Generation 1

SPEA2 combined archive and search populations, DTLZ2, 4-objectives



Generation 10

SPEA2 combined archive and search populations, DTLZ2, 4-objectives



Generation 100

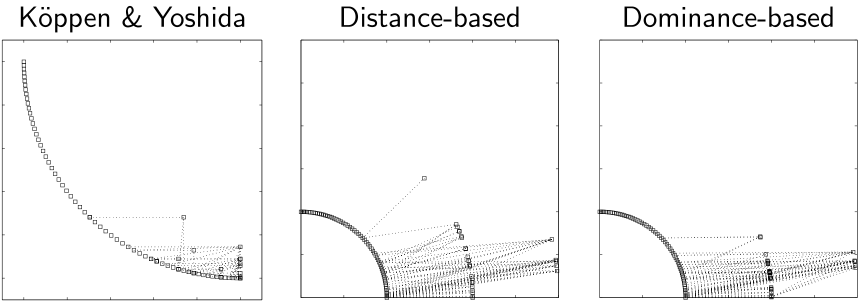
Points to consider and further work

- Spectral seriation gives good results empirically and is significantly faster than casting as an evolutionary optimisation problem
- The dominance-based visualisation is deterministic and can be plotted for a single pass through the data after seriation
  - Pareto shell membership preserved
  - Dominance preserved
  - Non-dominance not preserved (not possible in general), but misinformation minimised
- Areas to work on:
  - *Distance* between points in the original space not really conveyed
  - *Shape* of the fronts not conveyed
  - In *many*-objective problems the population rapidly becomes mutually non-dominated, point-based visualisations less useful

Thanks and questions...

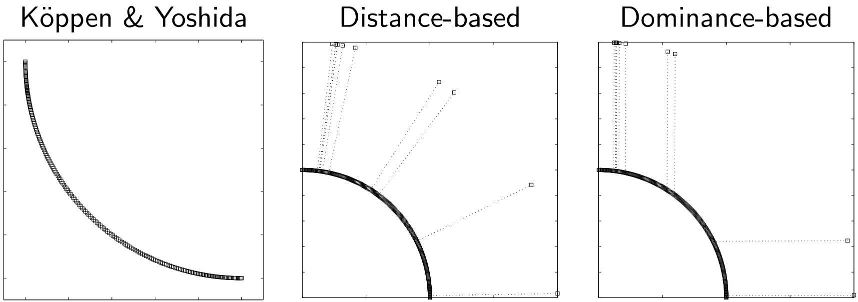
- Thank you
- Any questions?

SPEA2 combined archive and search populations, DTLZ2, 10-objectives



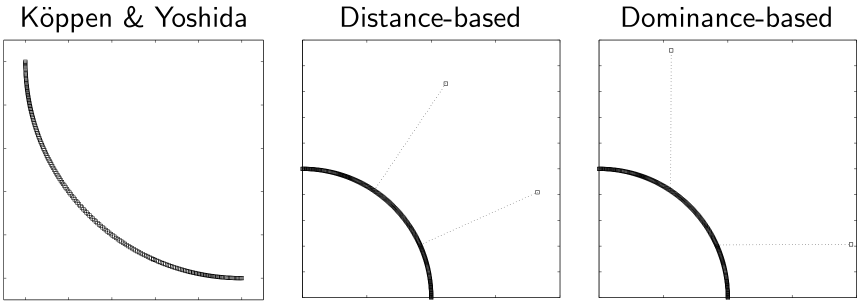
Generation 1

SPEA2 combined archive and search populations, DTLZ2, 10-objectives



Generation 10

SPEA2 combined archive and search populations, DTLZ2, 10-objectives

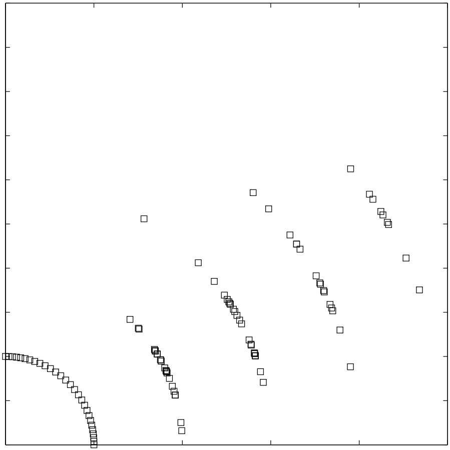


Generation 100

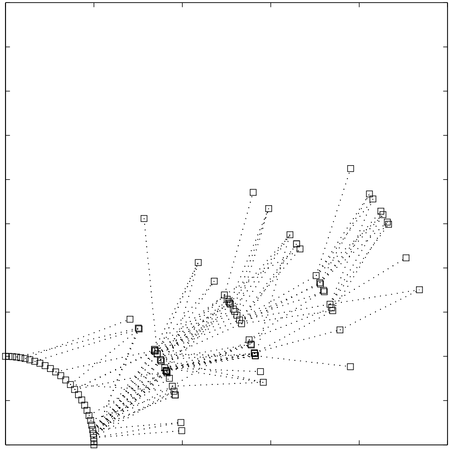
Distance-based

- The distance between shells in the mapping is arbitrary, so we use the angle of the ray passing through a mapped point and the origin to determine the placement of dominated solutions.
- Specifically, the location of a $\mathbf{u}_i$ is initially placed on the ray through the origin whose angle is the average of the angles of the rays associated with the mapped points which dominate it.
- As the position of $\mathcal{S}_0^2$ is determined using spectral seriation, the rays defining $\mathcal{S}_1^2$, can be rapidly computed, which, along with $\mathcal{S}_0^2$ can then be used to fix $\mathcal{S}_2^2$, and so on.
- An iterative procedure is used to adjust the locations of $\mathcal{S}_i^2$ points (where $i > 1$), such that the mean of the angles in $\mathbb{R}^2$ of those points which are *dominated* in $\mathbb{R}^D$, as well as those which dominate in $\mathbb{R}^D$, are used to set the location angles of $\mathcal{S}_i^2$ members.

Dominance denoted by *closeness*



Dominance denoted by *closeness*



Dominance denoted by *closeness*