The Future of Design of Experiments: is it "Optimal"?

Few insights on the applications and problems

Daniel Griffith & Eduardo Santiago

Minitab, Inc.

State College, PA

Minitab® Statistical Software

Outline

- Optimal Designs
- In what sense are optimal designs, optimal?
- When do we use optimal designs?
- A quick illustration of D-optimal designs
- Available Algorithms: Point-exchange, Coordinate-exchange, and genetic algorithms.
- Examples
- The future

Minitab® Statistical Software

Optimal Designs

Optimal designs are experiments that have been derived using a mathematical criterion. These criteria focus on different aspects of modeling, e.g. estimating the parameters of a linear model with minimum variance, or reducing the overall prediction variance across the design region.

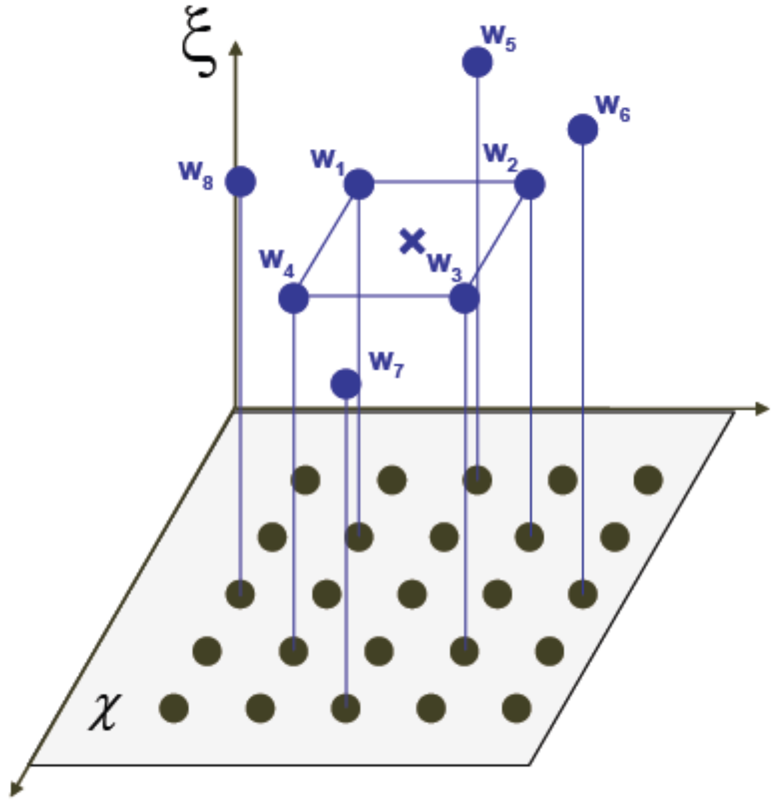


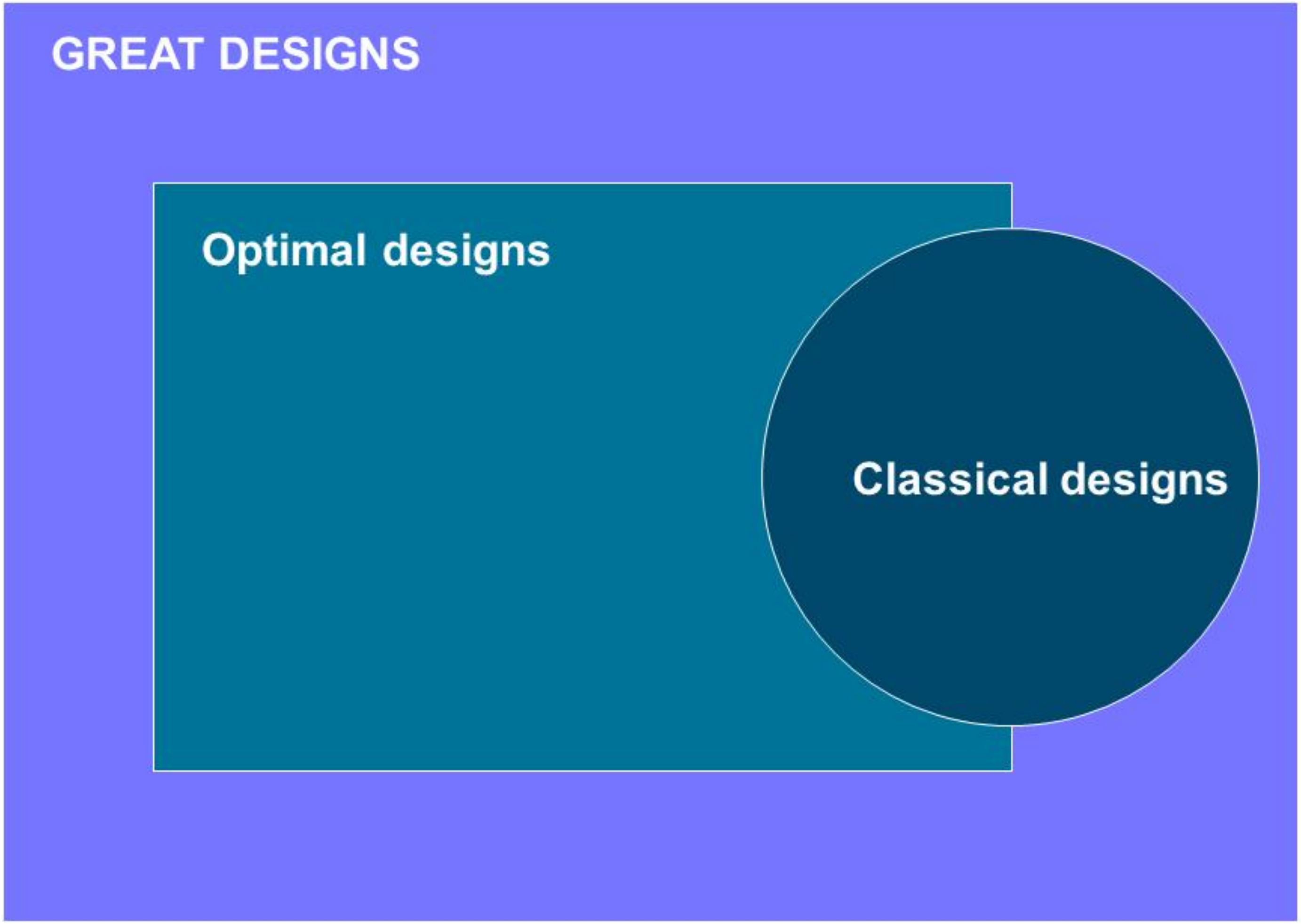
Fig 1. What is the best choice of runs for an experiment?

Optimal Designs

Optimal designs are **flexible** alternative approaches to the design of experiments. At the same time, optimal designs:

- Allow for the estimation of model parameters with *maximum precision*. This implies the confidence intervals for all coefficients will be minimized, which practically speaking means more certainty in the parameter estimates.
- *Small average prediction variance* across the design region. This property guarantees the prediction intervals will be narrow.

Optimal Designs



In what sense are optimal designs, optimal?

- Several optimality criteria exist (D-, I-, A-, E-, G-, etc.) however they are ALL related to the design in model form **(X),** the reason why is shown next.

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \text{where } \boldsymbol{\varepsilon} \sim \mathrm{N}\left(\mathbf{0}, \sigma^2\mathbf{I}\right)$$

$$\text{From OLS}: \ \hat{\boldsymbol{\beta}} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y}$$

$$\mathrm{Var}\left(\hat{\boldsymbol{\beta}}\right) = \sigma^2\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1} = f\left(\mathbf{X}\right)$$

Fisher's information matrix

- The two most predominant are D- and I-optimality.

In what sense are optimal designs, optimal?

- Some of these criteria (e.g. D-, E- and A-optimality) emphasize *maximum precision*, so the focus is on minimizing the inverse of the Fisher information matrix.
- While some of the other criteria focus on prediction variance (e.g., I- and G-optimality).
- Optimal designs are typically “optimal” only with respect to one of these criteria. [*Sometimes designs will be optimal with respect to multiple criteria, but this is rarely the case*]

A Quick Illustration – Power Wash Opt.

A team is investigating a power-wash process to clean steel couplings. Four important factors in the wash process and their current settings are considered. Suppose a Central Composite Design with 30 runs is proposed…**do we need all 30 experimental runs to arrive to similar conclusions?**

Variable	Description
Pressure (x_1)	Water pressure, psi (80, 140, 200)
Temp. (x_2)	Water temperature, ºF (160, 175, 190)
Time (x_3)	Wash cycle time, minutes (2, 4, 6)
Conc. (x_4)	Detergent concentration, % (2, 4.5, 7)

Minitab® Statistical Software

A Quick Illustration – Power Wash Opt.

- 30 runs provide sufficient information to estimate several terms in the model. In particular we can estimate: the constant, four main effects, six two-way interactions, four quadratic effects, and 15 DF allocated to estimate the error.
- Are we performing too many runs?
- Can we sacrifice a little orthogonality but save money by performing a shorter experiment?
- Will the quality of the proposed solution be the same with fewer runs?

Minitab® Statistical Software

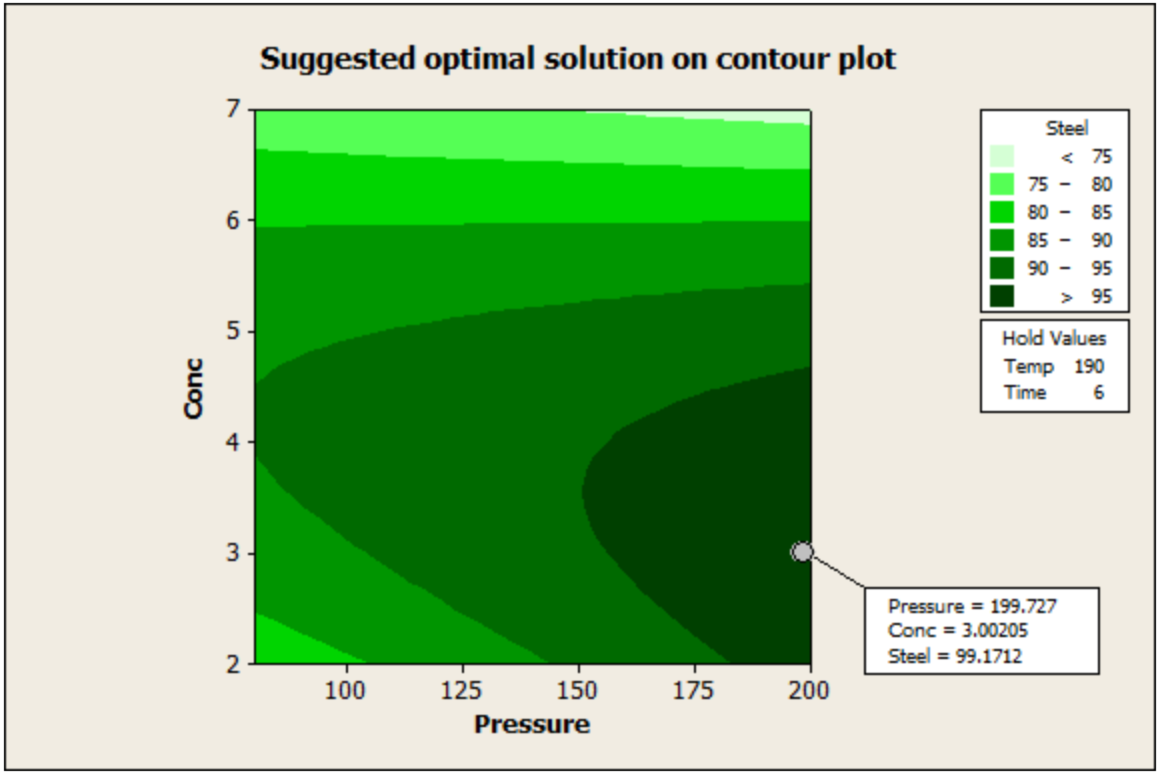
A Quick Illustration – Power Wash Opt.

The analysis below corresponds to the original 30 run CCD, using a significance level of 0.1.

```
The analysis was done using coded units.

Estimated Regression Coefficients for Steel

Term              Coef  SE Coef       T      P
Constant        89.217    1.867  47.789  0.000
Pressure         2.944    1.524   1.932  0.066
Temp             1.117    1.524   0.733  0.472
Time             2.622    1.524   1.720  0.099
Conc            -3.344    1.524  -2.194  0.039
Conc*Conc      -10.750    2.410  -4.460  0.000
Pressure*Conc   -4.700    1.617  -2.907  0.008
Temp*Conc       -3.900    1.617  -2.412  0.025


S = 6.46709    PRESS = 1568.41
R-Sq = 67.74%  R-Sq(pred) = 45.02%  R-Sq(adj) = 57.48%
```

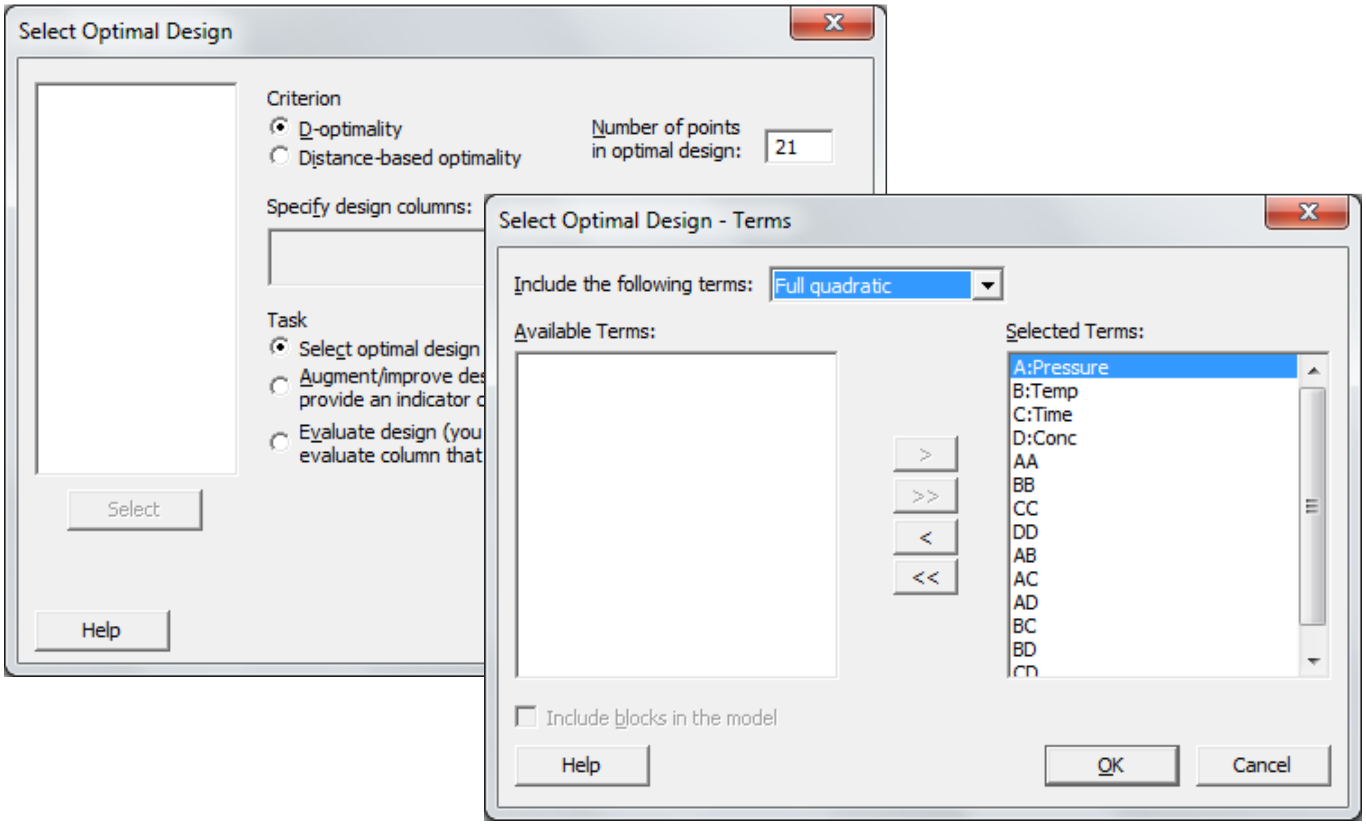
A Quick Illustration – Power Wash Opt.

The recommended solution is Pressure = 200 psi, Concentration = 3%. Temperature = 190, and Time = 6 seconds. Expected response = 99.17%.



A Quick Illustration – Power Wash Opt.

Would the conclusions have changed if you only select 21 of the original 30 runs? Is the optimal solution comparable to the obtained with the original design?



A Quick Illustration – Power Wash Opt.

Minitab proceeds to find the optimal subset of 21 runs that maximize the information about the model we want to fit.

PowerWs1.MTW ***

↓	C1	C2	C3	C4	C5	C6	C7	C8	C11
	StdOrder	RunOrder	PtType	Blocks	Pressure	Temp	Time	Conc	OptPoint
1	11	1	1	1	80	190	2	7.0	1
2	17	2	-1	1	80	175	4	4.5	0
3	16	3	1	1	200	190	6	7.0	1
4	30	4	0	1	140	175	4	4.5	0
5	22	5	-1	1	140	175	6	4.5	1
6	13	6	1	1	80	160	6	7.0	1
7	14	7	1	1	200	160	6	7.0	1
8	28	8	0	1	140	175	4	4.5	0
9	20	9	-1	1	140	190	4	4.5	1
10	10	10	1	1	200	160	2	7.0	1
11	25	11	0	1	140	175	4	4.5	0
12	29	12	0	1	140	175	4	4.5	0
13	26	13	0	1	140	175	4	4.5	0
14	5	14	1	1	80	160	6	2.0	1
15	15	15	1	1	80	190	6	7.0	1
16	4	16	1	1	200	190	2	2.0	1
17	9	17	1	1	80	160	2	7.0	1
18	2	18	1	1	200	160	2	2.0	1
19	19	19	-1	1	140	160	4	4.5	1
20	1	20	1	1	80	160	2	2.0	1
21	3	21	1	1	80	190	2	2.0	1
22	27	22	0	1	140	175	4	4.5	0
23	8	23	1	1	200	190	6	2.0	1
24	18	24	-1	1	200	175	4	4.5	1
25	12	25	1	1	200	190	2	7.0	1
26	21	26	-1	1	140	175	2	4.5	0
27	23	27	-1	1	140	175	4	2.0	1

Minitab® Statistical Software

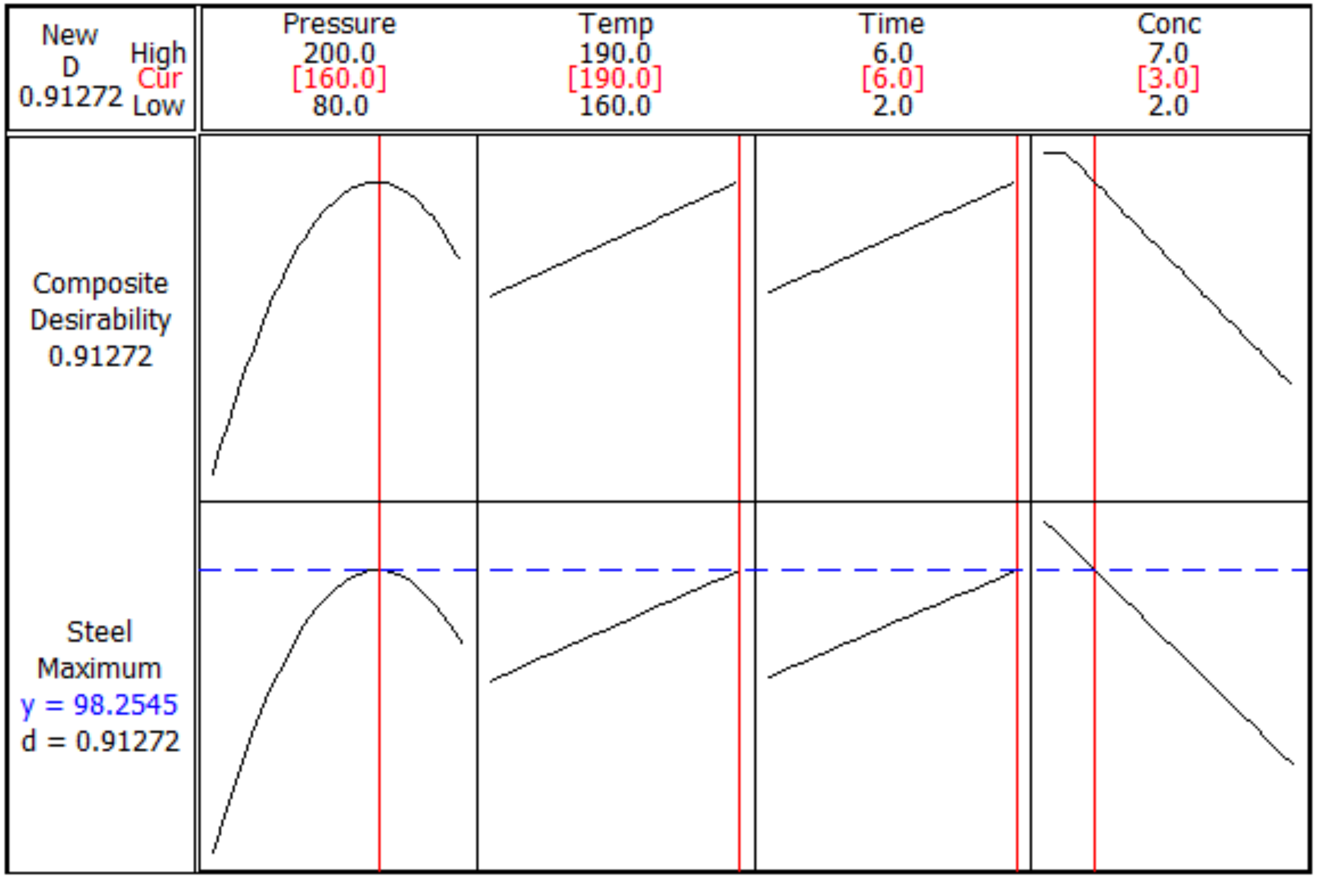
A Quick Illustration – Power Wash Opt.

Is the D-optimal design orthogonal? If not, how strongly correlated will the terms in the model be?

```
           Pressure      Temp      Time
Temp          0.000
              1.000

Time         -0.003     0.000
              0.990     1.000

Conc          0.003     0.000     0.003
              0.990     1.000     0.990


Cell Contents:  Pearson correlation
                P-Value
```

A Quick Illustration – Power Wash Opt.

Using the same responses for the complete CCD, we can fit a model to the corresponding responses collected for the 21-run D-optimal design.



Minitab® Statistical Software

When do we use optimal designs?

Classical designs not always apply.

- Design for nonstandard sample sizes
- Design in the presence of irregular design spaces
- Design for constrained mixture experiments
- Design augmentation (except if folding fractional design)
- Design for experiments having both qualitative and quantitative factors
- Design for nonlinear regression models
- Designs for response surface models with an assumed third order polynomial or higher

When do we use optimal designs?

- Unusual requirements concerning either the number of blocks or the block size
- Designs for logistic regression and other exponential family models
- Designs for situations in which there is heteroscedasticity across the design space
- Designs that model potential and active factors (supersaturated designs)

Minitab® Statistical Software

THE ALGORITHMS

Available algorithms

Several algorithms have been devised to create optimal designs. In general, these algorithms fall under one of the three categories mentioned below:

1. **Point-exchange algorithms** (e.g. Fedorov, DETMAX), require the user to specify a candidate set of points (N). These algorithms start with a random design with n runs. The algorithm then assesses which point (a.k.a. run) from the current design can be replaced with any of the candidate points to improve the optimality criterion.
2. **Coordinate-exchange algorithms**. Does not require a candidate set of points and evaluates the impact of changing a specific coordinate of the design at a time. Computationally faster.
3. **Genetic algorithms**. Flexible to find any criterion, perform a continuous search of the design space (no candidate set), but computationally slower than the algorithms falling under the other two categories. Does not require a random start design.

Coordinate Exchange Algorithm

Why is it so powerful? It's faster b/c computations are based on simpler changes on coordinates not row vectors (as traditionally done by point exchange methods).

Start by taking the number of factors, type, levels, runs, and assumed linear model. Create a random design, such that |**X'X**| ≠ 0. For example, for a four run design with two continuous factors and a main effects model:

We find the best change in [-1, 1] to maximize the determinant

$$\mathbf{X} = \begin{bmatrix} 1 & -0.53 & -0.66 \\ 1 & -0.67 & 0.57 \\ 1 & 0.63 & 0.94 \\ 1 & 0.4 & -0.08 \end{bmatrix}$$

For D-optimality:

|**X'X**| = 6.43039

Minitab® Statistical Software

Coordinate Exchange Algorithm

The trick to the efficiency of the algorithm is that there is a computationally cheap formula ([illegible]) to calculate the change in the determinant corresponding to the change in that specific coordinate.

X_{12}	[illegible]	|X'X|
-0.53	1	6.4304
-1	1.2036	7.7398
0	0.9022	5.8017
1	1.0984	7.0633

Best change is –1; now proceed to find the best coordinate exchange for –0.67.

$$\mathbf{X} = \begin{bmatrix} 1 & -1 & -0.66 \\ 1 & -0.67 & 0.57 \\ 1 & 0.63 & 0.94 \\ 1 & 0.4 & -0.08 \end{bmatrix}$$

Coordinate Exchange Algorithm

- You iterate multiple times over every coordinate until the change in the determinant is minimum or |**X'X**| reaches its upper bound of n^p.
- The order in which the columns are optimized can be selected at random.

EXAMPLES

Design Augmentation

A manufacturer of roofing shingles wants to know the key variables that affect granule loss – the amount of granules removed from a shingle during a scrub test. Engineers have suggested that six variables could significantly impact granule loss.

Variable	Description
Moisture (x_1)	Amount of moisture on shingles (0.5, 1.5)
Pressure (x_2)	Application pressure, psi (20, 36)
Sheet Tension (x_3)	Sheet tension in pounds (10, 35)
Viscosity (x_4)	Coating Viscosity, centipoise (2000, 4500)
Filler%. (x_5)	Percentage of filler used in coating (40, 50)
Line Speed (x_6)	Speed of the line in ft/min (650, 750)

Design Augmentation

After analyzing and reducing the model…

Estimated Effects and Coefficients for GranuleLoss (coded units)

Term	Effect	Coef	SE Coef	T	P
Constant		244.69	4.172	58.65	0.000
Press	47.12	23.56	4.172	5.65	0.000
Viscosity	-71.13	-35.56	4.172	-8.52	0.000
Filler%	-47.38	-23.69	4.172	-5.68	0.000
LineSpeed	-21.88	-10.94	4.172	-2.62	0.026
Press*Viscosity	-24.88	-12.44	4.172	-2.98	0.014

S = 16.6872   PRESS = 7128.64
R-Sq = 93.85%   R-Sq(pred) = 84.25%   R-Sq(adj) = 90.77%

Design Augmentation

```
Factors:    6   Base Design:         6, 16   Resolution:   IV
Runs:      16   Replicates:              1   Fraction:    1/4
Blocks:     1   Center pts (total):      0
```

```
Design Generators: E = ABC, F = BCD
```

```
Alias Structure

I + ABCE + ADEF + BCDF

A + BCE + DEF + ABCDF
B + ACE + CDF + ABDEF
C + ABE + BDF + ACDEF
D + AEF + BCF + ABCDE
E + ABC + ADF + BCDEF
F + ADE + BCD + ABCEF
AB + CE + ACDF + BDEF
AC + BE + ABDF + CDEF
AD + EF + ABCF + BCDE
AE + BC + DF + ABCDEF
AF + DE + ABCD + BCEF
BD + CF + ABEF + ACDE
BF + CD + ABDE + ACEF
ABD + ACF + BEF + CDE
ABF + ACD + BDE + CEF
```

Since this is a Resolution IV design, what if the engineers suspect that the CF interaction could have been significant? How do you estimate both terms?

Design Augmentation

Block	Moisture	Pressure	SheetT	Viscosity	Filler%	LineSpeed
1	0.5	20	10	4500	40	750
1	1.5	36	35	2000	50	650
1	1.5	36	10	2000	40	750
1	0.5	20	35	4500	50	650
1	0.5	36	35	2000	40	650
1	0.5	36	10	4500	50	650
1	0.5	36	35	4500	40	750
1	1.5	20	10	2000	50	650
1	1.5	20	35	2000	40	750
1	1.5	20	35	4500	40	650
1	1.5	36	10	4500	40	650
1	0.5	20	35	2000	50	750
1	0.5	36	10	2000	50	750
1	0.5	20	10	2000	40	650
1	1.5	36	35	4500	50	750
1	1.5	20	10	4500	50	750
2	1.5	36	35	4500	40	650
2	0.5	36	10	2000	40	650
2	0.5	20	10	2000	50	750
2	1.5	20	35	4500	50	750

Using the coordinate exchange algorithm, we can find four additional runs that allow for the estimation of both interactions. D-efficiency is 89%.

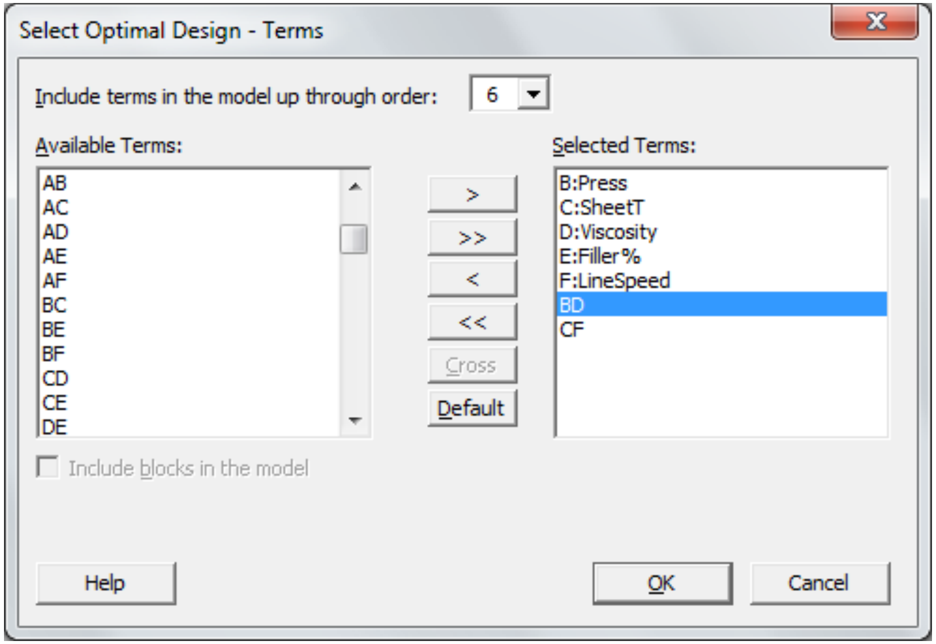
Design Augmentation

	Moisture	Pressure	SheetT	Viscosity	Filler%	LineSpeed	BD
Pressure	0.000						
SheetT	0.200	0.000					
Viscosity	0.200	0.000	0.200				
Filler%	0.000	-0.200	0.000	0.000			
LineSpeed	0.000	-0.200	0.000	0.000	0.200		
BD	0.000	0.000	0.000	0.000	0.000	0.000	
CF	0.000	0.000	0.000	0.000	0.000	0.000	0.600

Cell Contents: Pearson correlation

Design Augmentation

- Folding is usually an option. That would add 16 more runs to the experiment. However, folding a Resolution IV design does not guarantee dealiasing of the confounded terms.
- If you could have anticipated that these two interactions would be significant, you could have chosen a design with 20 runs that minimizes the partial confounding and allows for the estimation of all terms on the right.



Design Augmentation

- Minitab's D-optimal design for 20 runs shows a low degree of multicollinearity between the terms of interest.

	Press	SheetT	Viscosity	Filler%	LineSpeed	BD
SheetT	-0.200					
Viscosity	0.000	0.000				
Filler%	-0.000	0.000	-0.000			
LineSpeed	0.000	0.000	0.200	0.000		
BD	0.000	0.000	0.000	0.000	0.000	
CF	0.000	0.000	0.000	0.000	0.000	-0.200

Cell Contents: Pearson correlation

- Sometimes however, you can't anticipate the importance of terms that can be confounded with each other.

Balanced Incomplete Block Designs

Four catalysts are being investigated in an experiment. The experimental procedure consists of selecting one of the six batches of raw material, loading the pilot plant, applying each catalyst in a separate run and observing the reaction time. The batches of raw material are considered as blocks, however each batch is only large enough to permit three catalysts to be run.

- One factor with four treatments.
- Six blocks with a limitation that forces only three of the treatments to be tested under each block.
- N = 18 runs total are needed.

Balanced Incomplete Block Designs

Use your code's interface and solve this problem. Make a picture with your design like the one below. BIBD for one factor can be determined empirically when you consider only one factor, but why is this approach advantages? Designs can be created for any type of blocking (not just balanced).

	block											
treatment	1	2	3	4	5	6						
A	A	A	A	-	-	-	1	1	1	0	0	0
B	B	-	-	B	B	-	1	0	0	1	1	0
C	-	C	-	C	-	C	0	1	0	1	0	1
D	-	-	D	-	D	D	0	0	1	0	1	1

$$a = 4, b = 6, k = 2, r = 3, \lambda = 1, \mathcal{N} = (n_{ij})_{4\times 6}$$

Disadvantages of using an OD

- The model is assumed to be known. However, classical designs also assume that! Screening experiments assume only main effects are significant; response surface designs assume only quadratics, main effects, and two-way interactions are significant.
- Treating it like a black box optimization routine can produce designs that are poor choices. (Example)
- Optimization problem is difficult to solve and no current algorithm will give the absolute globally best design. It is always good to check if running the command multiple times generates the same design (software should start at different random starting points every time).

The Future

- Most practitioners find Design of Experiments overwhelming, especially when it to picking the right design.
- Optimal design routines can make that easier and more transparent.
- However it is necessary to add controls and checks to see if a classical design with center points might be a suitable choice as well, instead of just going with the I- or D-optimal design.

Minitab® Statistical Software

APPENDIX

Appendix: Optimality Criteria

- The *ultimate goal* is to produce a design that minimizes the uncertainty in the parameter estimates.
- This is equivalent to minimizing the variance of the estimated coefficients.

$$\mathrm{Var}\left(\hat{\boldsymbol{\beta}}\right) = \sigma^2\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1} = f\left(\mathbf{X}\right)$$

Minitab® Statistical Software

The D-optimality criterion

- *D-optimality* is probably the most common criterion and mathematically it maximizes the determinant of ($\mathbf{X}^T\mathbf{X}$) which is equivalent to minimizing the variances and covariances in the equation above.
- This optimality criterion is measured in multiple ways:

$$D_{\text{eff}} = \frac{\left|\mathbf{X}^T\mathbf{X}\right|^{1/p}}{N}$$

SAS and JMP report this number.

$$D_{\text{eff}} = \left(\frac{\left|\mathbf{X}^T\mathbf{X}\right|}{\max\limits_{\zeta^*} \left|\zeta^* \zeta\right|} \right)^{1/p}$$

The A-optimality criterion

- *A-optimality* focuses on minimizing the sum of the coefficient variances. This criterion results in minimizing the average variance of the estimates of the regression coefficients.

$$\min_{\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_N} \; tr\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}$$

- There is another criterion related to Fisher's information matrix, this is E-optimality which attempts to minimize the maximum eigenvalue of the inverse of $\mathbf{X}^{\mathrm{T}}\mathbf{X}$.

Minitab® Statistical Software

Criteria based on prediction variance

- There are other criteria associated with the scaled prediction variance, *v(**x**)*.

$$v(\mathbf{x}) = \mathrm{N} \cdot \mathbf{x}^{\mathbf{T}} \left(\mathbf{X}^{\mathbf{T}}\mathbf{X}\right)^{-1} \mathbf{x}$$

- *G-optimality* focuses in minimizing the maximum prediction variance over the points chosen in the design space.

$$\max_{\mathbf{x}} \; \mathrm{N}\mathbf{x}^{\mathbf{T}} \left(\mathbf{X}^{\mathbf{T}}\mathbf{X}\right)^{-1} \mathbf{x} \ge p \quad \Rightarrow \quad \mathrm{G}_{\mathrm{eff}} = \frac{p}{\max_{\mathbf{x}} \; \mathrm{N}\mathbf{x}^{\mathbf{T}} \left(\mathbf{X}^{\mathbf{T}}\mathbf{X}\right)^{-1} \mathbf{x}}$$

Criteria based on prediction variance

- *I-optimality* is the integrated variance criterion and it minimizes the average variance prediction over a continuous region and not just *m* distinct points.

$$\min_{\mathbf{X}} \; \frac{1}{\int_{\mathrm{R}} d\mathbf{x}} \times \int_{\mathrm{R}} \mathbf{x}^{\mathrm{T}} \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1} \mathbf{x} \, d\mathbf{x}$$

Minitab® Statistical Software