

Multivariat optimering

Kurs i Bergen 1998-10-30

Tomas Öberg

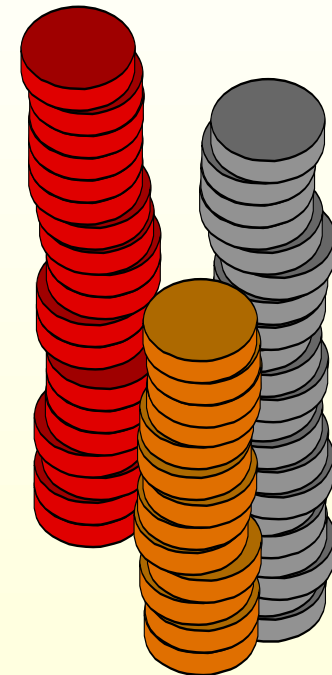
Innehåll

- Varför optimera?
- Vilka förberedelser är nödvändiga.
- Olika sätt att hantera flera optimeringsmål.
- Systematiska optimeringsmetoder: EVIT, gradient, simplex och responsystemetodik.
- ”Evolutionär tillverkning” (EVOP)
- Automatisk optimering.
- Sammanfattning.

Optimering lönar sig!

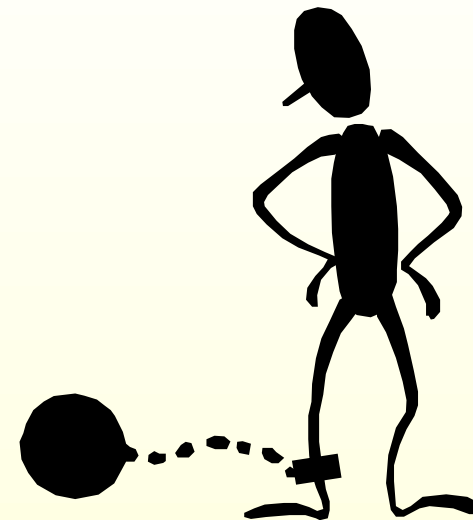
”In a typical manufacturing plant, for example, increased productivity, improved product quality, lowered operating costs can pay back millions of dollars to users over a short time.”

Automation Research Corporation 1997



”Låsta rattar”

- Trots det kör många fortfarande med ”låsta rattar”. Varför?
- Kanske är det föreställningen om den ideala processen, som aldrig förändrar sig. Men finns den?
- Är det kanske komplexiteten? Är det för svårt?



Visst är det komplicerat

- Flera optimeringsmål måste beaktas samtidigt.
- Flera experimentvariabler har betydelse.
- Randvillkoren för det experimentella området ("domänen") är svåra att ange.
- Det "optimala" utgångsläget för optimeringsarbetet är okänt.



Men inte olösligt!

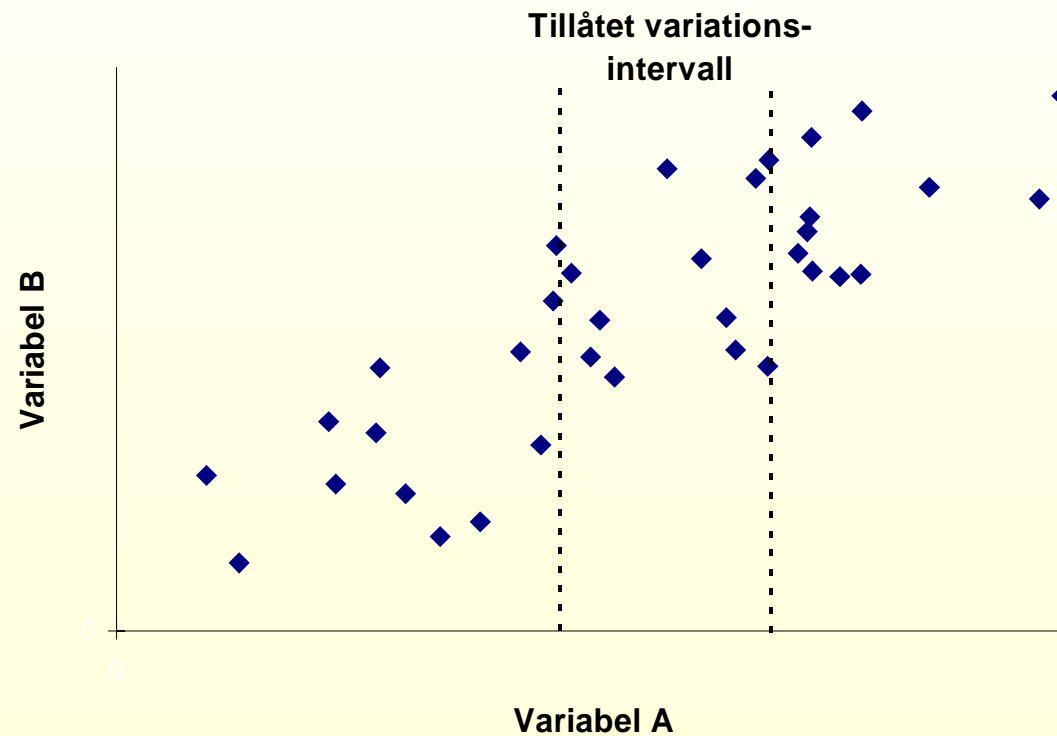
- Flera optimeringsmål kan hanteras samtidigt.
- Och vi kan studera flera experimentvariabler på samma gång.
- Randvillkoren för det experimentella området behöver inte definieras först.
- Vi söker reda på det ”optimala” utgångsläget.



Varför experimentell optimering?

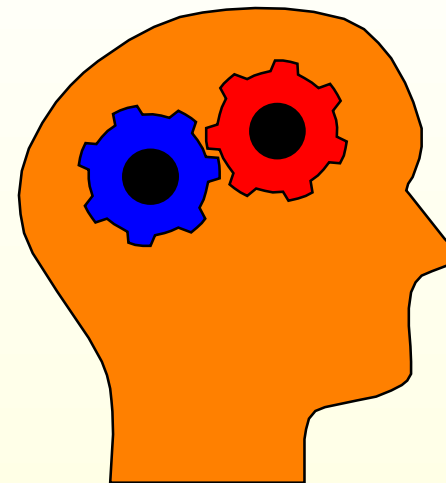
- Optimering är att justera ett system för att uppnå bästa möjliga resultat ("respons").
- Experimentell optimering innebär att optimeringen baseras på praktiska försök.
 - De flesta tekniska, och kemiska, system kan inte beskrivas matematiskt med tillräcklig precision.
 - Numerisk optimering är därför inte möjlig.

Data och information är inte samma sak



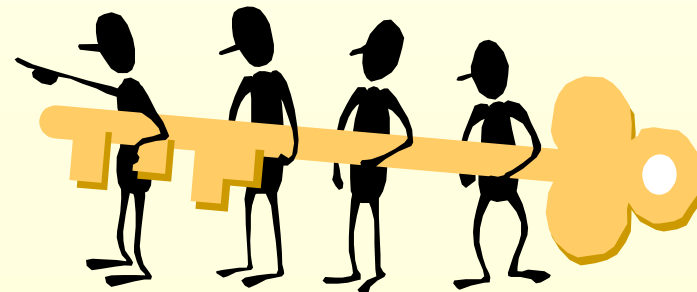
Nödvändiga förberedelser

- Bestäm optimeringsmålen.
 - Ofta en balansgång.
 - Process-/produktkunskap.
 - Consensus.
- Välj ut experimentvariablerna.
 - Lista allt som kan inverka.
 - Välj ut det som är påverkbart.
 - Prioritera utifrån bakgrundkunskap.
 - > 10 st, genomför ”screening”-försök.



Consensus-mötet

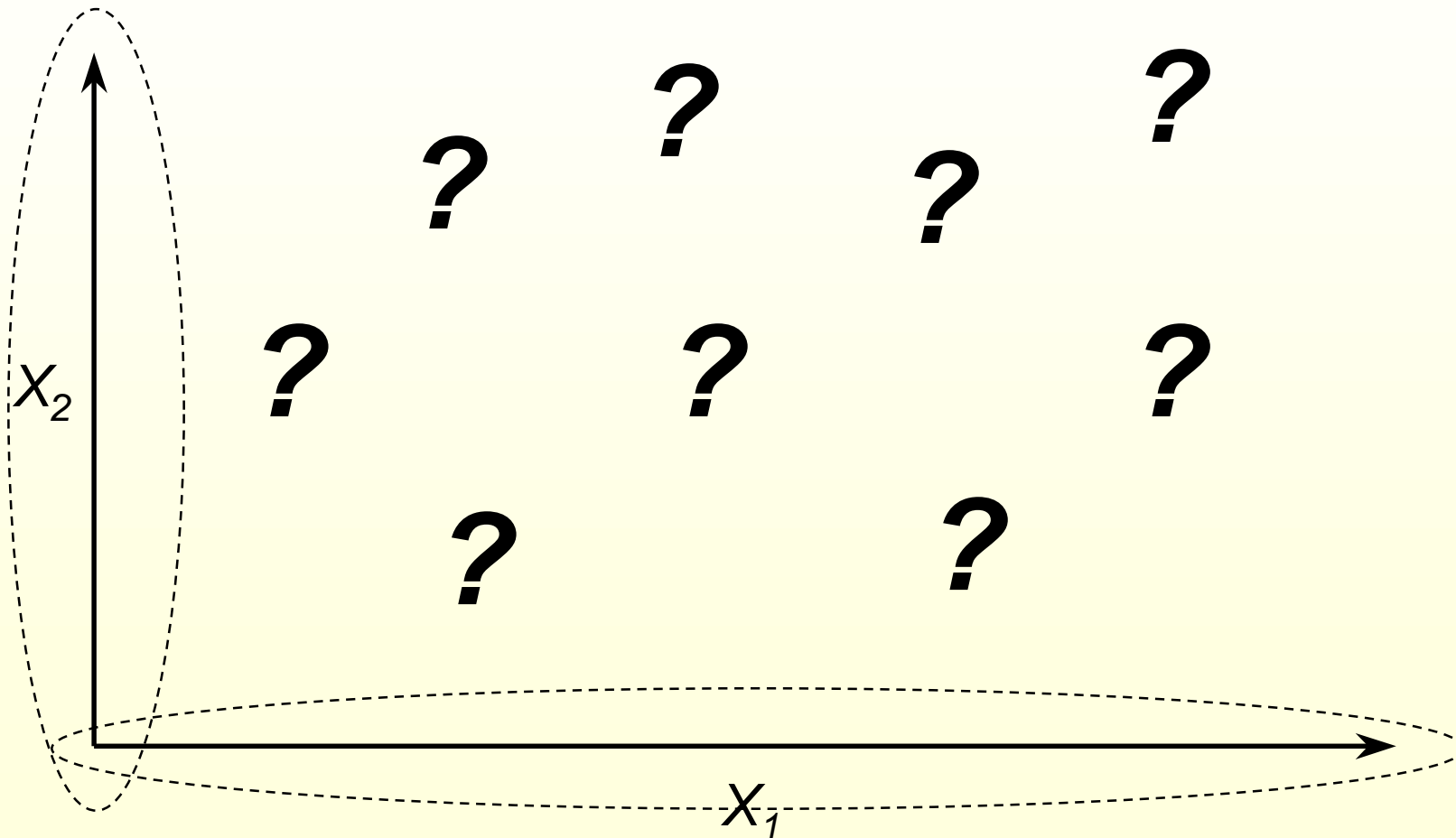
- Representanter för alla inblandade funktioner (marknad/försäljning, utveckling och produktion).
- Alla blir delaktiga och motiverade.
- Resultaten blir accepterade.



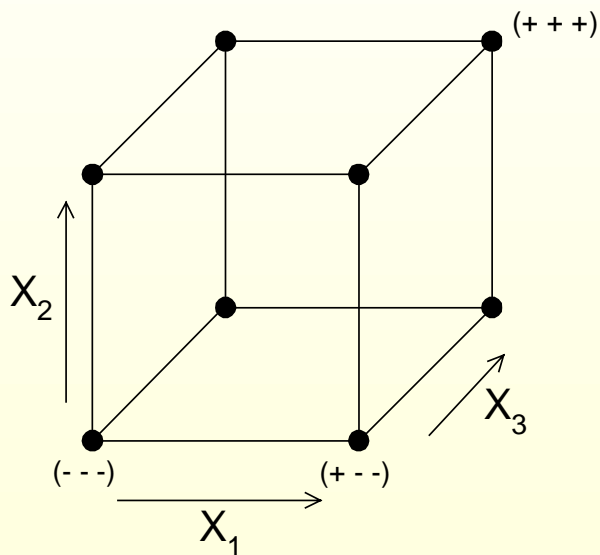
”Screening”-försök

- Motiverat om vi har fler än 10 st potentiellt viktiga och påverkbara experiment-variabler.
- Syftet är att prioritera för att nå en påtaglig förbättring snabbare.
- Screening-försöken ska inte utföras genom att variera en variabel i taget (EVIT).
- Lämpliga försöksplaner är s.k. reducerade faktorförsök och Plackett-Burman försök

EVIT ("en variabel i taget")



Faktorförsök



exp #	X_1	X_2	X_3
2	-	-	-
6	-	-	+
5	-	+	-
8	-	+	+
4	+	-	-
1	+	-	+
3	+	+	-
7	+	+	+

Utvärdering av faktorförsök

Beräkning av en effekt:

Multiplitera tecknet i avsedd kolumn med responsvärdet, addera, och dela med antalet experiment.

A	B	C	AB	AC	BC	ABC	Y
-	-	-	+	+	+	-	121.7
+	-	-	-	-	+	+	123.5
-	+	-	-	+	-	+	125.2
+	+	-	+	-	-	-	122.8
-	-	+	+	-	-	+	124.5
+	-	+	-	+	-	-	122.1
-	+	+	-	-	+	-	127.2
+	+	+	+	+	+	+	124.8

Utvärdering av ett faktorförsök innebär anpassning av en polynommodell som även kan innefatta termer för att beskriva samspel.

$$y = b_0 + \sum b_i x_i + \sum \sum b_{ij} x_i x_j + e$$

Variansanalys

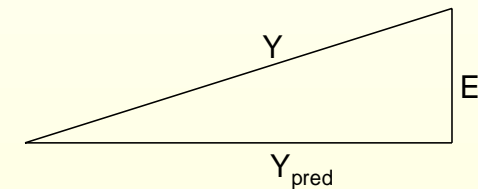
Anta en linjär modell: $y = b_0 + b_1 \cdot A + b_2 \cdot B + b_3 \cdot C$

$b_0 = 123.975$, $b_1 = -0.675$, $b_2 = 1.025$ och $b_3 = 0.675$.

Y	Y-Y _m	(Y-Y _m) ²	Y _p	Y _p -Y _m	(Y _p -Y _m) ²	Y-Y _p	(Y-Y _p) ²
121.7	-2.27	5.18	122.95	-1.03	1.05	-1.25	1.56
123.5	-0.47	0.23	121.60	-2.38	5.64	1.90	3.61
125.2	1.23	1.50	125.00	1.03	1.05	0.20	0.04
122.8	-1.18	1.38	123.65	-0.32	0.11	-0.85	0.72
124.5	0.53	0.28	124.30	0.32	0.11	0.20	0.04
122.1	-1.88	3.52	122.95	-1.03	1.05	-0.85	0.72
127.2	3.23	10.40	126.35	2.38	5.64	0.85	0.72
124.8	0.83	0.68	125.00	1.03	1.05	-0.20	0.04
		23.16			15.70		7.46
LINJÄR MODELL	fg	KvS	MKv	F	p-värde		
Regression	3	15.70	5.23	2.81	0.17		
Residualer	4	7.46	1.87				
Totalt	8	23.16					

Pytagoras sats:

$$Y^2 = Y_{pred}^2 + E^2$$



Kvoten jämförs med det kritiska värdet i F-fördelningen med givet antal frihetsgrader och vald signifikansnivå: $F_{3,4, p < 0.05} \Rightarrow 6.59$. Modellen är alltså inte statistiskt signifikant.

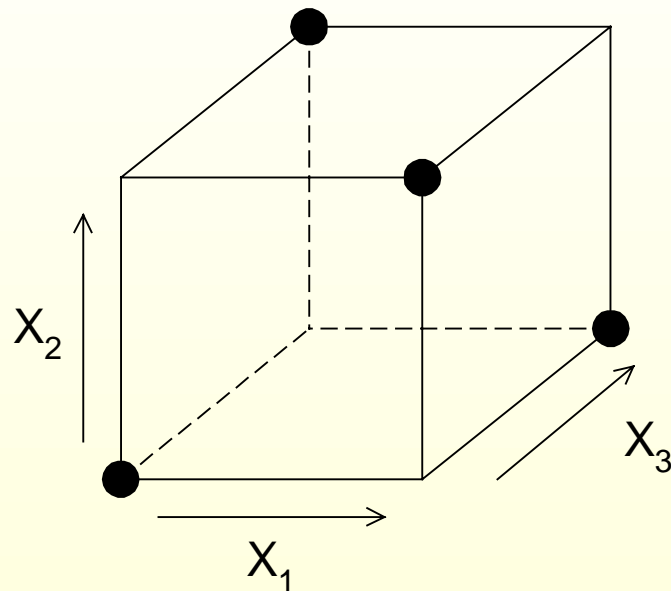
Residualspridningen (icke-beräknade effekter) kan också användas för att skatta osäkerheten i beräkningen av koefficienterna: s^2 för koefficienterna = $1.87/8 = 0.23 \Rightarrow s = 0.48$. K.I. = $t_{0.05} \cdot s = 2.8 \cdot 0.48 = 1.34$.

Bristande modellanpassning

- Definieras som den del av standardavvikelsen för residualerna som beror på faktorer som inte beskrivs av modellen.
- I "mättade" försöksplaner, antal försök = antal modelltermer + 1, kan man inte beräkna residualvariansen.
- För att särskilja experimentfelet behövs även replikat:
Residualvarians = modellfel + experimentfel

$$F = \frac{S_{\text{mod}}^2}{S_{\text{exp}}^2}$$

Reducerade faktorförsök



X_1	X_2	X_3
a	b	-(ab)
-	-	-
-	+	+
+	-	+
+	+	-

Hur många försök behövs i en
”screening”?

Var.	Försök	Var	Försök
8-11	12	20-23	24
12-15	16	24-27	28
16-19	20	28-31	32

Tolkning av effekter

- Fokusera på att använda beräknade effekter som ett hjälpmedel att prioritera.
- Det går även att beräkna statistisk signifikans.
 - Förutsätter att antalet beräknade effekter är mindre än antalet försök minus ett.
 - Förutsätter att vi vet hur modellen ser ut (linjär, interaktion, kvadrat, osv.)
 - Kom-ihåg att det är helt olika saker att påvisa att statistisk signifikans finns och att det saknas!

Egenskaper hos responsvariablerna

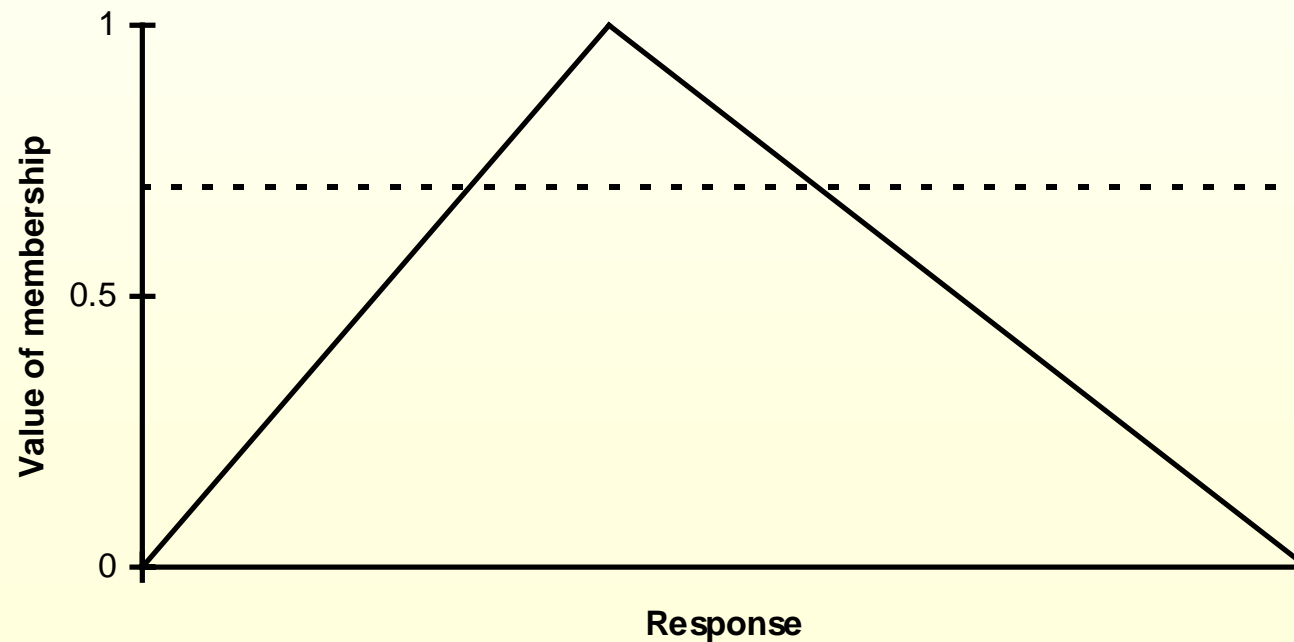
- Definierar optimeringsmålen:
 - Maximering.
 - Minimering.
 - Målvärde.
- De mäts med olika skalor.
- Den relativa betydelsen ("vikten") är inte densamma.

Hantera flera optimeringsmål

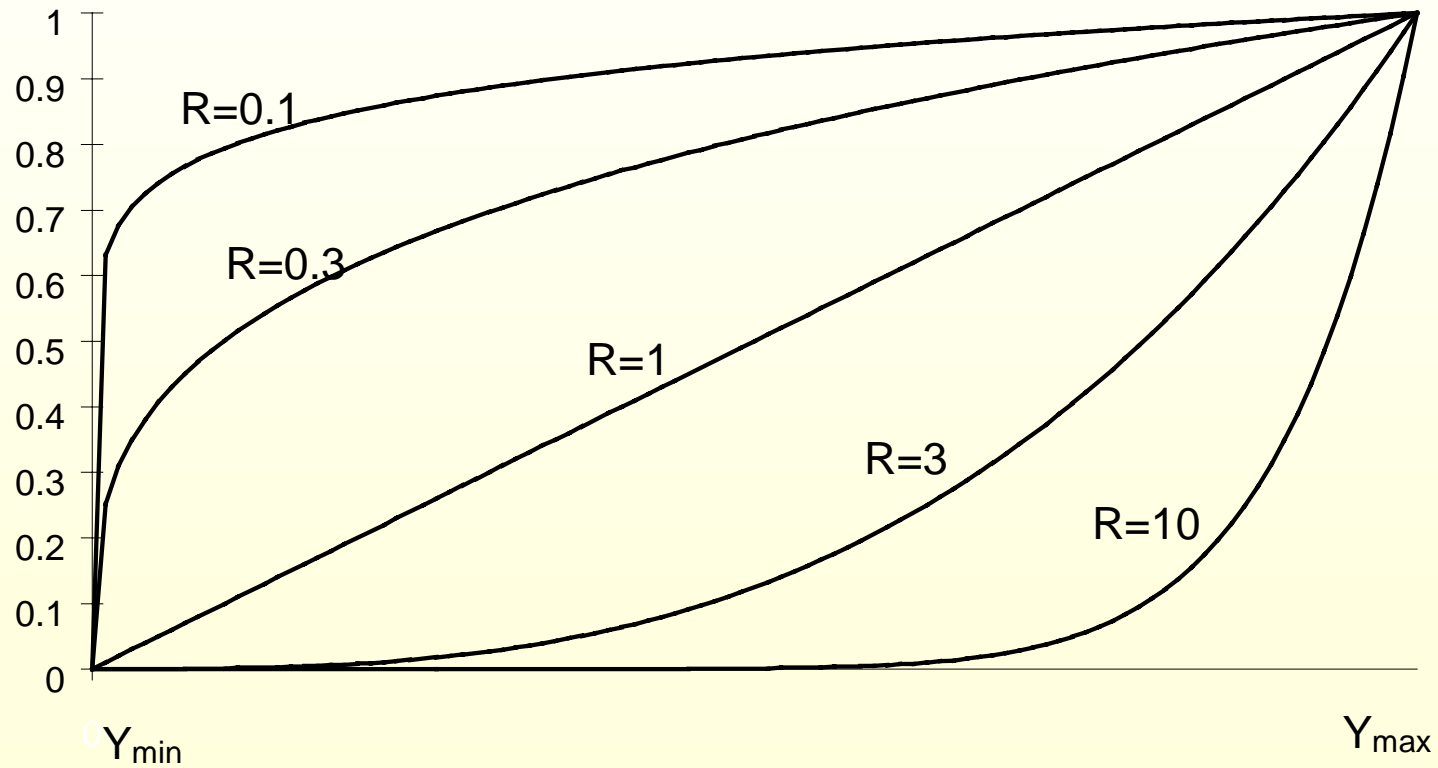
- “Objektiva” funktioner
 - Exempelvis förhållandet mellan upplösning och separationstid.
 - Leder ofta till optimering mot enbart det ena målet.
- Tillhörighetsfunktioner (“desirability”)
 - Avvägning mellan många olika mål.
 - Det mest populära tillvägagångssättet idag.
- Konturplottar ovanpå varandra
 - Max två experimentvariabler åt gången.
 - Blir ofta röriga (helst ej mer än två responsvariabler.)

En tillhörighetsfunktion

Optimeringsmålen representeras med funktioner som varierar mellan 0 och 1.



Olika mål - olika former



Sammanvägd tillhörighet

$$M(y) = \frac{\sum_{i=1}^I \beta_i \cdot m(y_i)}{\sum_{i=1}^I \beta_i}, \text{ if all } m(y_i) > 0$$

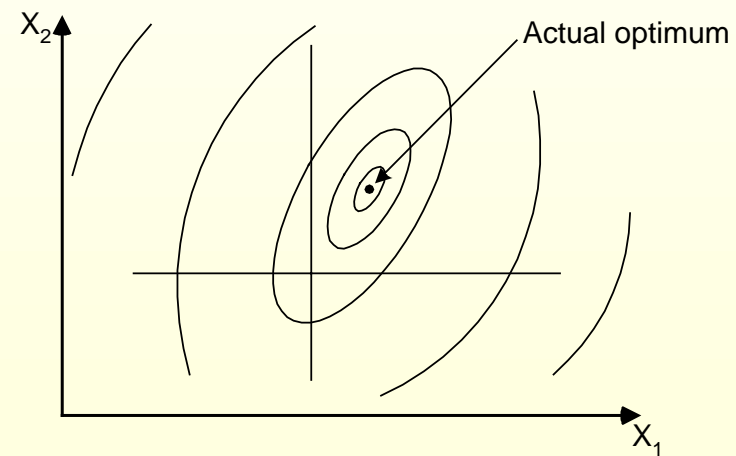
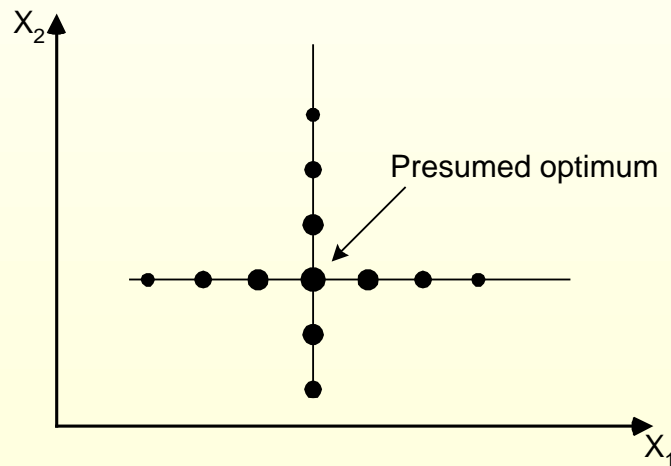
$$M(y) = 0, \text{ if any } m(y_i) = 0$$

- $M(y)$ = sammanvägd tillhörighet.
- $m(y_i)$ = tillhörighet för de individuella responsvariablerna.
- β_i = inflytande/vikt för de individuella responsvariablerna.
- I = antalet responsvariabler.

Systematiska optimeringsmetoder

- ”En-variabel-i-taget” (EVIT).
- Sekventiella försöksplaner, simplex-och gradientmetoden.
- Simultana försöksplaner, responsytemetodik.

”En-variabel-i-taget” (EVIT)



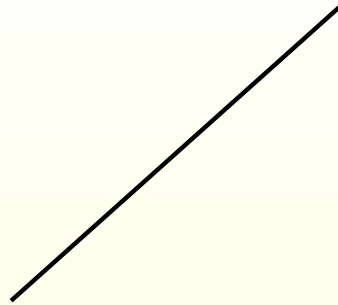
Simplex-metoden

- Enkel att förstå, och enkel att använda.
- Först $k+1$ försök, en "simplex".
- Därefter genomförs försök i den riktning som ger en förbättring.
- Utmärkande för metoden:
 - Sekventiell (steg-för-steg).
 - Inga antaganden om modell (linjär, kvadratisk, osv).
 - Kostnadseffektiv, få försök jämfört med responsytemetodik.

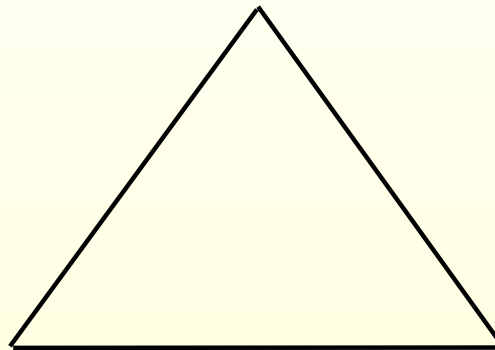
Historik

- **1962**, den grundläggande simplex-metoden presenteras av Spendley, Hext & Himsworth.
- **1965**, den modifierade simplex-metoden presenteras av Nelder & Mead.
- **1975**, anpassning av den modifierade simplex-metoden till praktisk experimentell verksamhet av Deming m.fl.
- **1980-talet**, olika ”nya” simplex-metoder (”super-modified”, ”composite modified”, etc.)

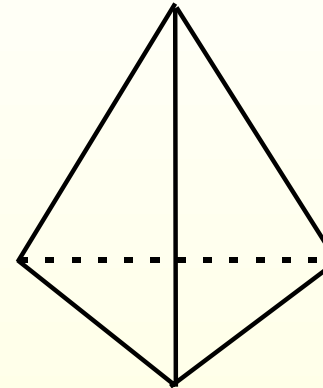
Vad är en simplex?



Linje



Triangel



Tetraeder

Startförsöken

- Klassisk/tilted (Spendley et al 1962).
- Cornered (Koshal)
- Optimal (linjär D-optimal, Öberg 1998)

$$D_0 = \begin{bmatrix} 000\dots 0 \\ pqq\dots q \\ qpq\dots q \\ \dots\dots\dots \\ qqq\dots p \end{bmatrix}$$

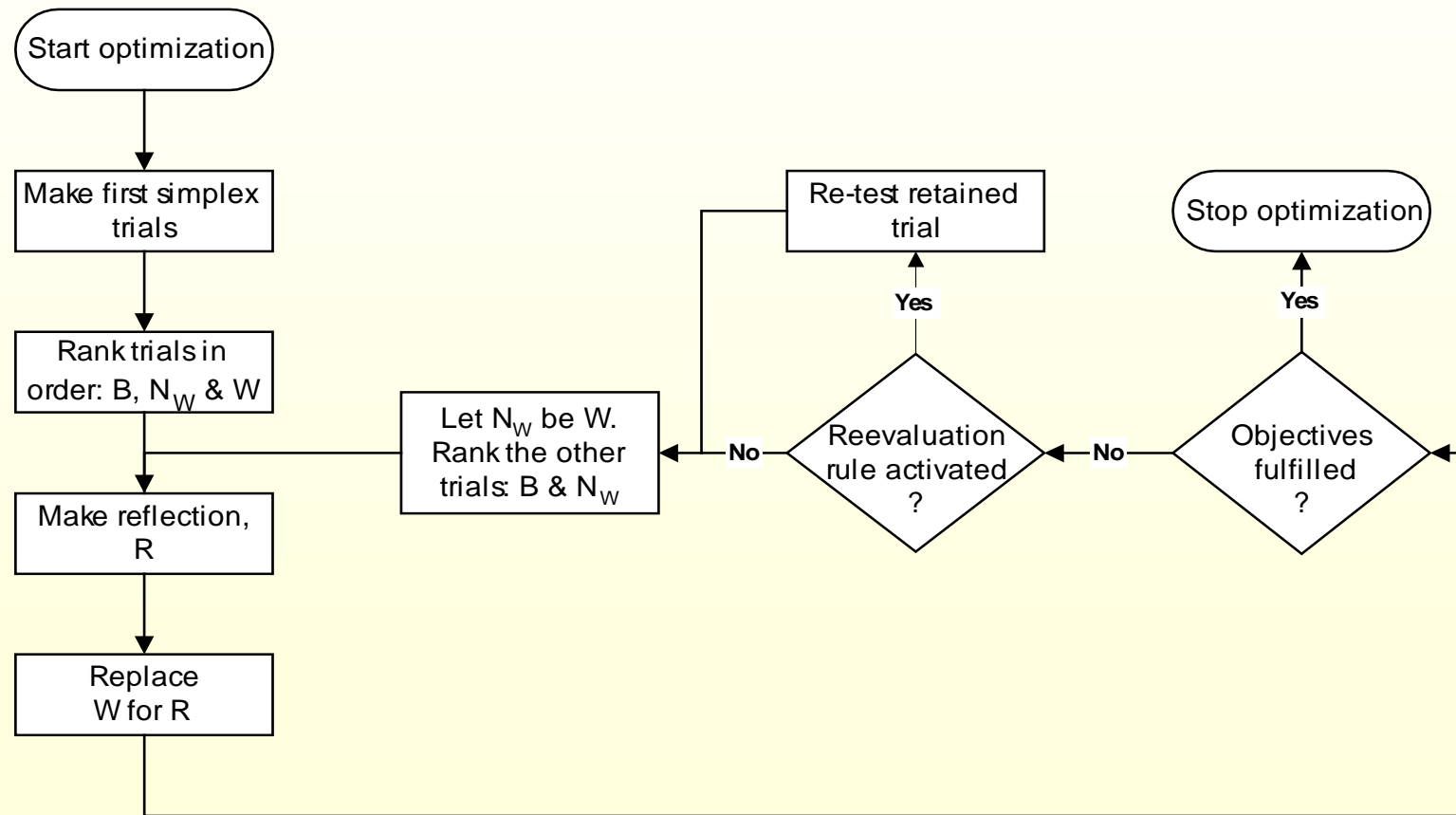
$$p = \frac{1}{k\sqrt{2}} \left\{ (k-1) + \sqrt{k+1} \right\}$$

$$q = \frac{1}{k\sqrt{2}} \left\{ \sqrt{k+1} - 1 \right\}$$

Grundläggande räkneregler

- Genomför $k+1$ försök.
- Rangordna försöken från det bästa (B) till det sämsta (W).
- Det sämsta provet utgår och nya inställningar beräknas genom spegling i centroiden (medelvärde) för de återstående proven.
- Återvänd ej till utgångsläget om det nya provet skulle bli det sämsta. Spegla istället det näst sämsta provet (Nw).
- Om ett prov har funnits med länge i simplexen ($k+1$ ggr) så görs försöket om.
- Prov utanför experimentdomänen ansätts ett oändligt dåligt värde.

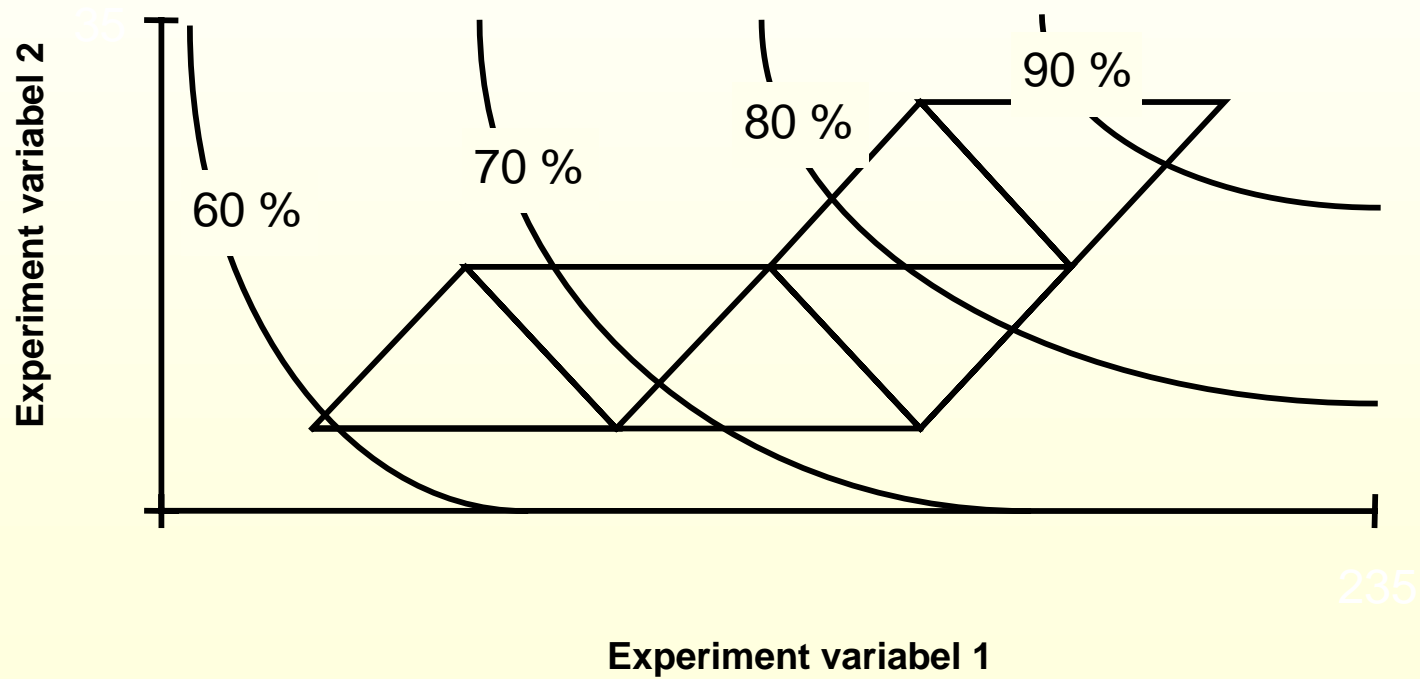
Flödesschema för den grundläggande metoden



Kalkylexempel

<i>Rank</i>	<i>Var 1</i>	<i>Var 2</i>
<i>Nw</i>	37.60	65.90
<i>B</i>	51.75	51.75
C (medelv.)	44.68	58.83
W	20.18	34.32
C-W	24.50	24.51
R=C+(C-W)	69.17	83.33

Steg för steg mot målet

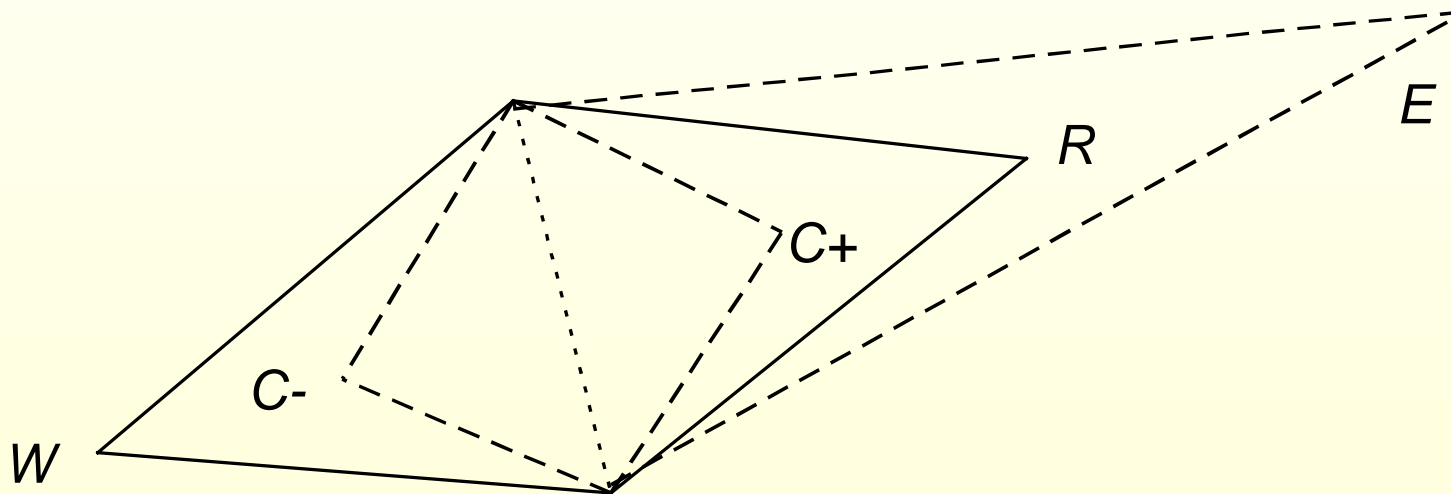


K+1-regeln är viktig

- Felaktigt dåliga respons är inget problem eftersom simplexen snabbt letar sig i rätt riktning igen.
- Felaktigt bra respons är ett större problem då simplexen fastnar.
- Försök som varit kvar en längre tid, $k+1$ eller $k+3$ ggr, måste därför utvärderas igen.
- Alternativ:
 - Ersätt det gamla resultatet med det nya.
 - Ta medelvärdet av gammalt och nytt.

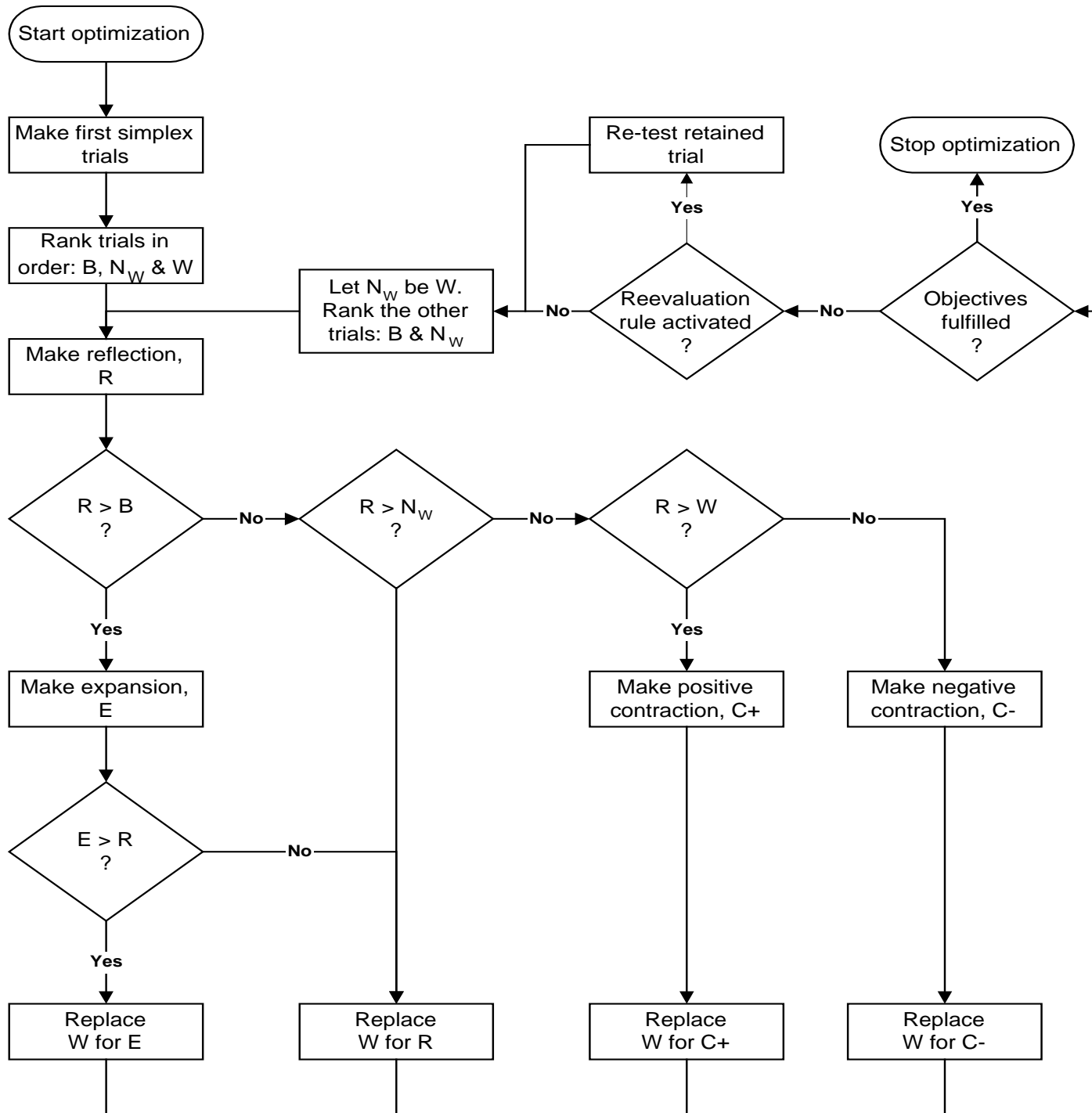
Modificerade räkneregler

- Simplexerna tillåts expandera i en riktning som innebär förbättring och dra sig samman i en riktning som innebär en försämring.



Modifierade räkneregler forts.

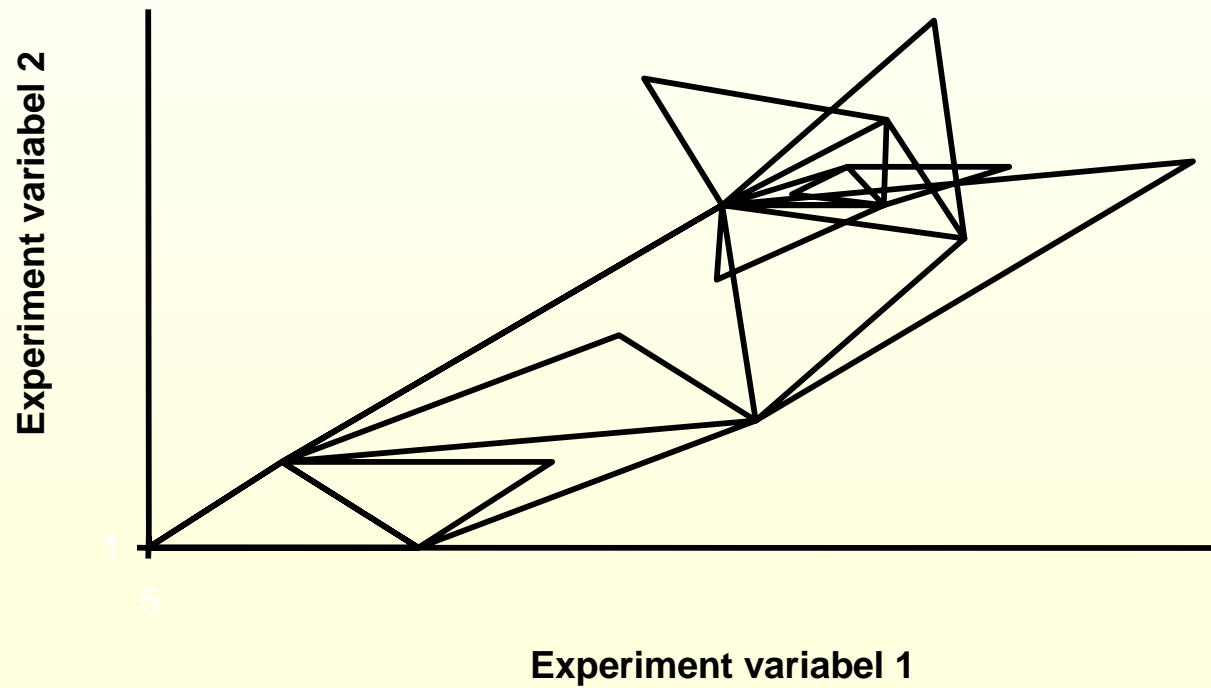
- Utvärdera R efter genomförd spegling.
 - Om $N_w < R = < B$, använd simplex $B..N_wR$.
 - Om $R > B$, beräkna och utvärdera E .
 - Om $E > R$, använd simplex $B..N_wE$.
 - Om $E = < R$, använd simplex $B..N_wR$.
 - Om $R = < N_w$:
 - Om $R > W$, beräkna och utvärdera $C+$, använd simplex $B..N_wC+$.
 - Om $R = < W$, beräkna och utvärdera $C-$, använd simplex $B..N_wC-$.



Kalkylexempel

Rank	Var 1	Var 2
Nw	37.60	65.90
B	51.75	51.75
C (medelv.)	44.68	58.83
W	20.18	34.32
C-W	24.50	24.51
R=C+(C-W)	69.17	83.33
(C-W)/2	12.25	12.25
C-=C-(C-W)/2	32.43	46.57
C+=C+(C-W)/2	56.92	71.08
E=R+(C-W)	93.67	107.84

Snabbt mot målet



Flexibilitet

- Lägg till, byt ut mot, egna försök.
- Lägg till experimentvariabler.
- Dra bort experimentvariabler.

Viktigt att kontrollera att simplexen inte degenererar, d.v.s. att den fortfarande kan söka i alla riktningar. Samtliga variabler i den aktuella simplexen måste därför vara linjärt oberoende.

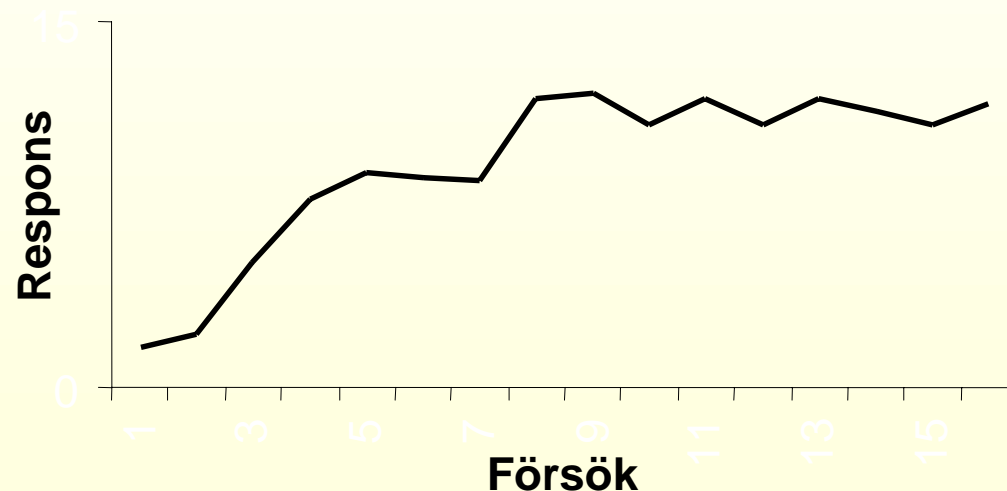
Speciella frågeställningar

- Multipla optima
 - Mycket ovanligt i den praktiska verkligheten.
 - Starta om optimeringen från flera olika punkter.
- Diskreta variabler/kvalitativa
 - Principiellt måste varje inställning optimeras var för sig.
 - Ibland kan diskreta variabler göras kontinuerliga, t.ex. 10 st spjäll som kan vara öppna eller stängda kan behandlas som en variabel (antal öppna spjäll).

När ska man sluta

- När man är nöjd.
- När ingen påtaglig förändring sker i respons och/eller de experimentvariabler som har haft betydelse.

Matematiska konvergensvillkor är däremot av tveksamt värde vid experimentell tillämpning.



Optimering av blandningar

- Många produkter är, eller består av, blandningar.
Exempelvis: Färger, lim, tvättmedel och läkemedel.
- Blandningsvillkoret försvårar ofta försöksplanering och utvärdering. $\sum x_i = 100\%$
- Simplexmetoden är lätt att tillämpa:
 - Alt. 1 Den sista variabeln beräknas som skillnaden mellan de övriga och 100%, och brott mot blandningsvillkoret behandlas som ett omöjligt försök.
 - Alt. 2 En fullständig och degenererad startsimplex anges inom blandningsdomänen.

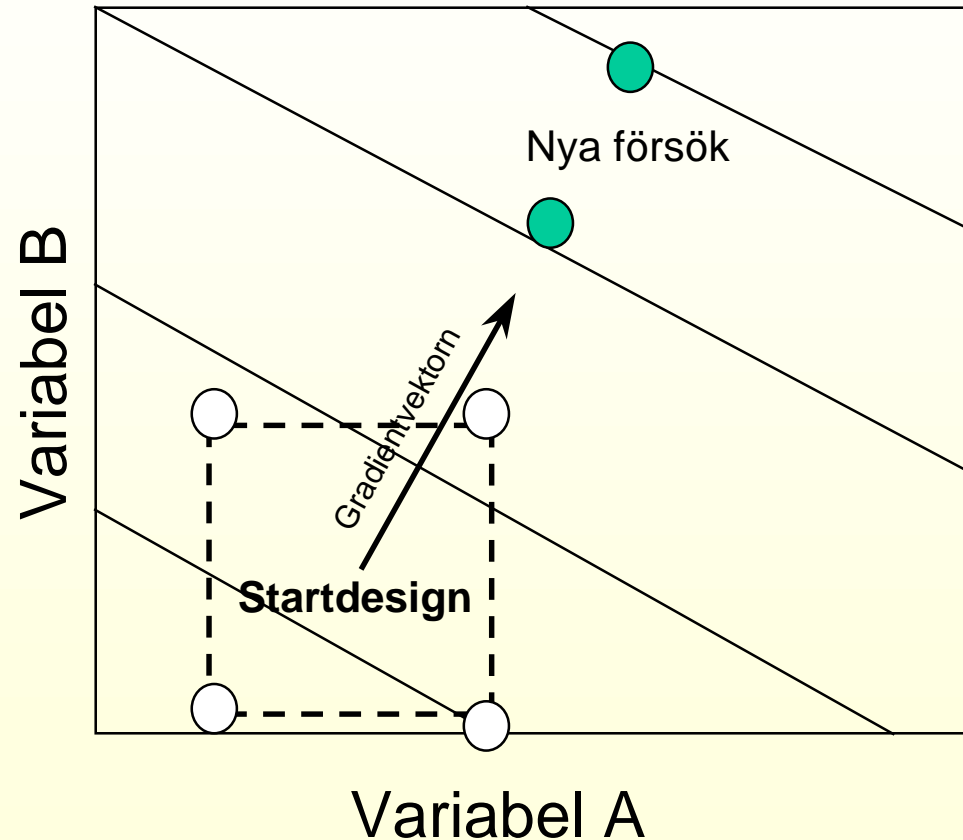
Gradientmetoden

Huvudeffekterna skattas med ett reducerat faktorförsök.

Optimering sker genom förflyttning längs gradientvektorn för snabbast ökning eller minskning.

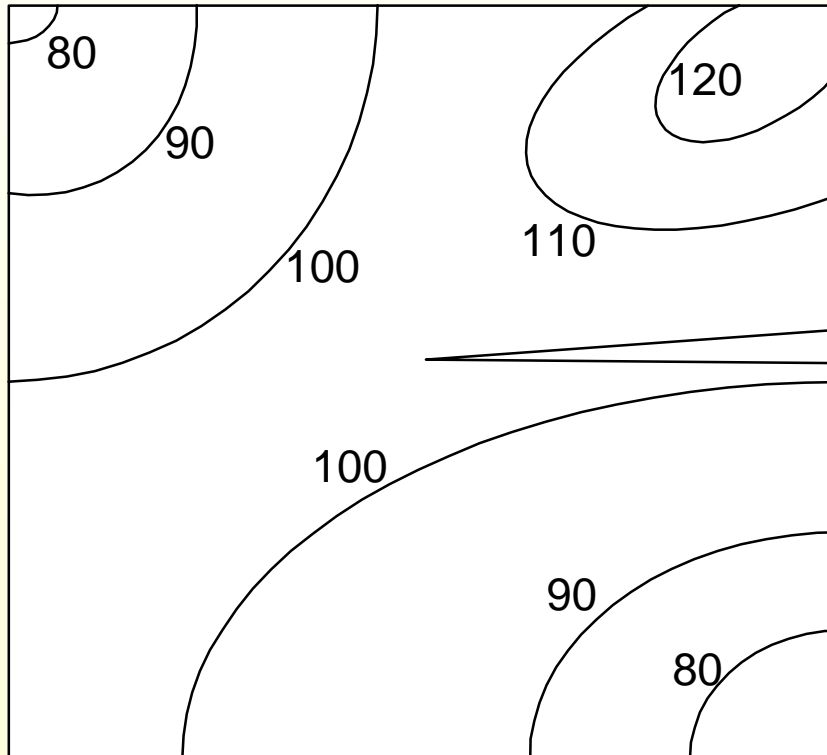
Gradientvektorn definieras av koefficienterna i det anpassade linjära polynomet:
 $y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 \dots b_nx_n$.

Förutsätter att huvudeffekterna dominerar och fungerar alltså inte ensam nära ett optimum.



Responssystemmodeller

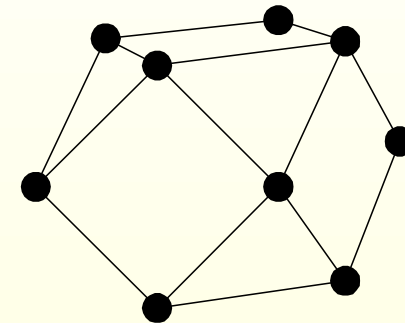
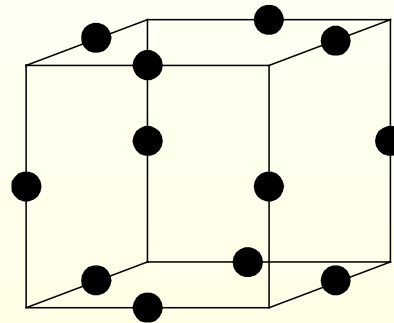
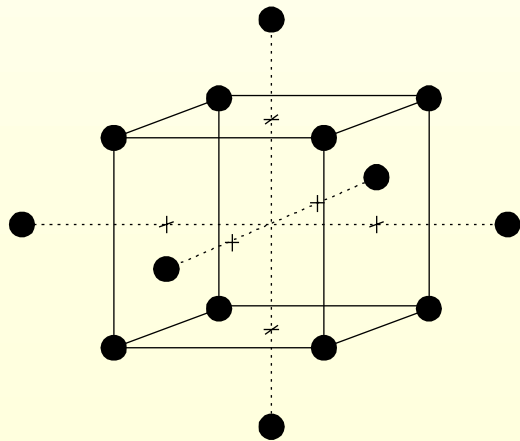
- Approximerar det verkliga sambandet med en kvadratisk modell: $y = b_0 + \sum b_i x_i + \sum b_{ii} x_i^2 + \sum \sum b_{ij} x_i x_j + e$
- Experiment-data anpassas till modellen.
- Används för att:
 - Tolka variabel-sambanden.
 - Hitta lämpligt driftsområde.



Driftsområde:
Område med
acceptabla värden
och liten variation.

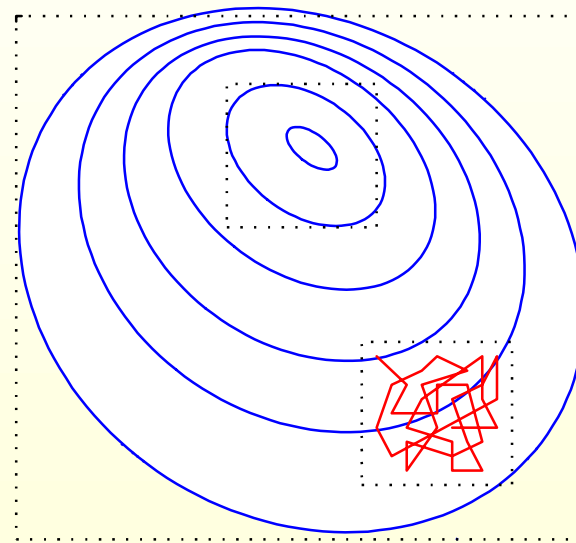
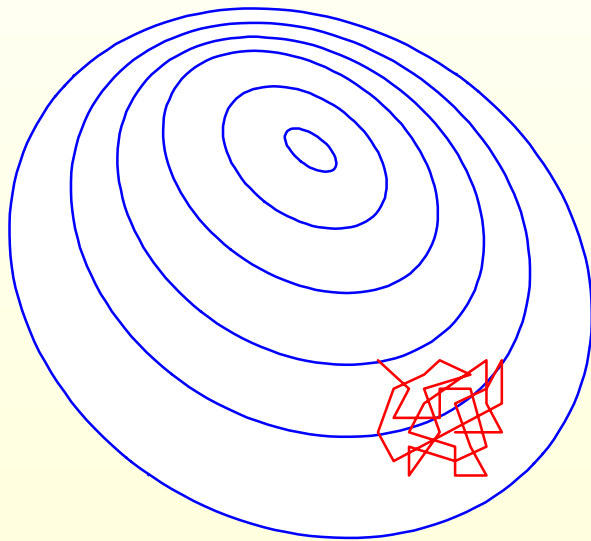
Simultana försöksplaner för att beräkna responsystemodeller (RSM)

Försök på minst tre nivåer.



<i>No. variables</i>	3	4	5	6
Box-Behnken	15	27	46	54
Central Composite	23	36	59	100

Var ska försöksplanen placeras?



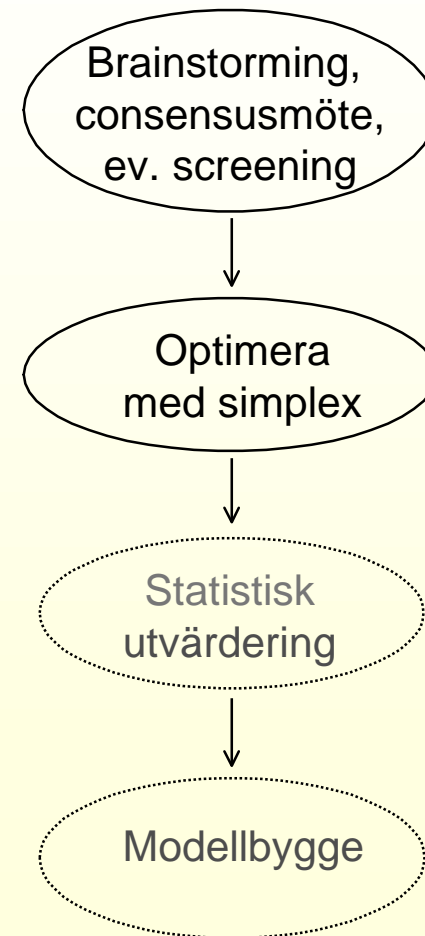
Vad är det vi vill ha svar på?

- Vilka variabler är viktiga?
- Hur ser funktionssambanden ut?
- Var är optimum?

Är det bara den sista frågan som ska besvaras så ska vi inte börja med de två första! Och vill vi ha svar på alla så ska vi gå i omvänd ordning.

En sammansatt strategi

- Steg ett. Optimering med simplex.
därefter
- Steg två. Utvärdera:
 - Effekterna av experiment-variablerna.
 - Statistisk signifikans.
 - Bygg modeller (t.ex. RSM).



Ett applikationsexempel från analytisk kemi

- PCB bestämdes med HRGC/ECD.
- Viktigt att hålla nere analys tiden, utan att äventyra nödvändig kromatografisk upplösning.
- Simplexoptimering av separationsbetingelserna är en rättfram lösning.

Experimentella förutsättningar

- Försök med Aroclor 1248, med PCB #209 som intern standard.
- 3 μ l, 10 ng PCB/ μ l, injicerades i "split mode".
- Kapillär: 30 m J&W DB5, 1.0 μ m film, 320 μ m i.d.
- Kromatograf: HP 5890, Series II med autoinjektor HP 7673.
- Detektor: ECD.

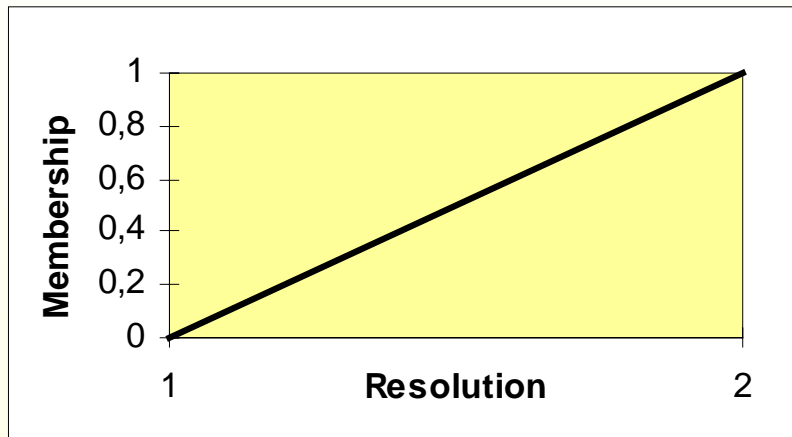
Inställningar av experiment-variablerna

	Ref. value	Step size
Initial temperature °C	190	10
Initial time min	4	2
Temperature gradient °C/min	6	2
Maximum oven temp. °C	300	10

Responsvariablerna

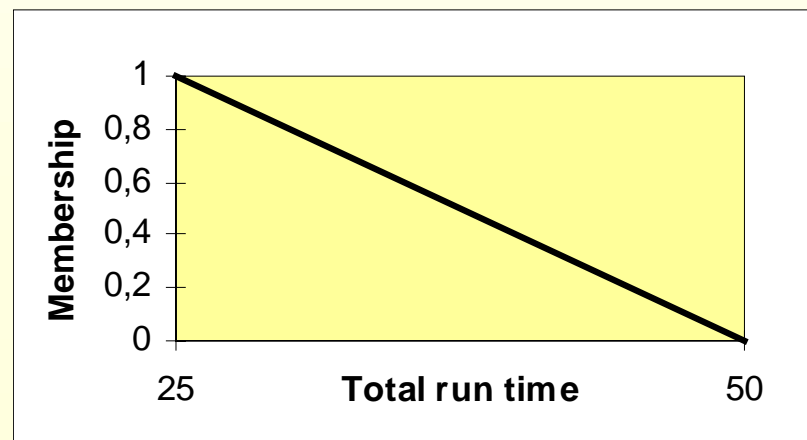
- Kromatografisk upplösning R1, mellan två toppar i början av kromatogrammet. $[R = \frac{X_A - X_B}{0.5(W_A + W_B)}]$
- Kromatografisk upplösning R2, mellan två toppar i den senare delen av kromatogrammet.
- Total separationstid.

Medlemsfunktionerna

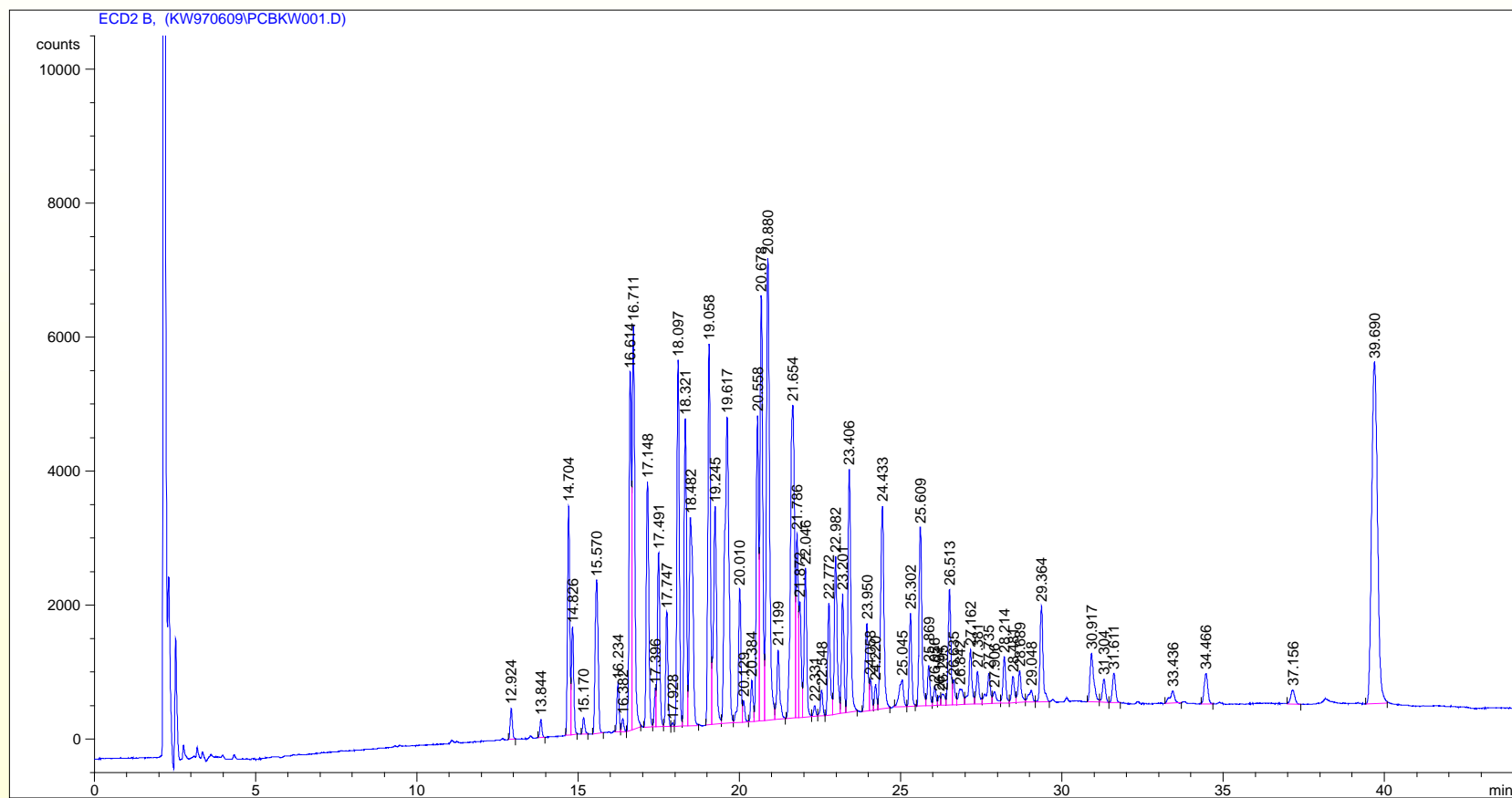


Influence = 0.67

Influence = 1



Kromatogram 1



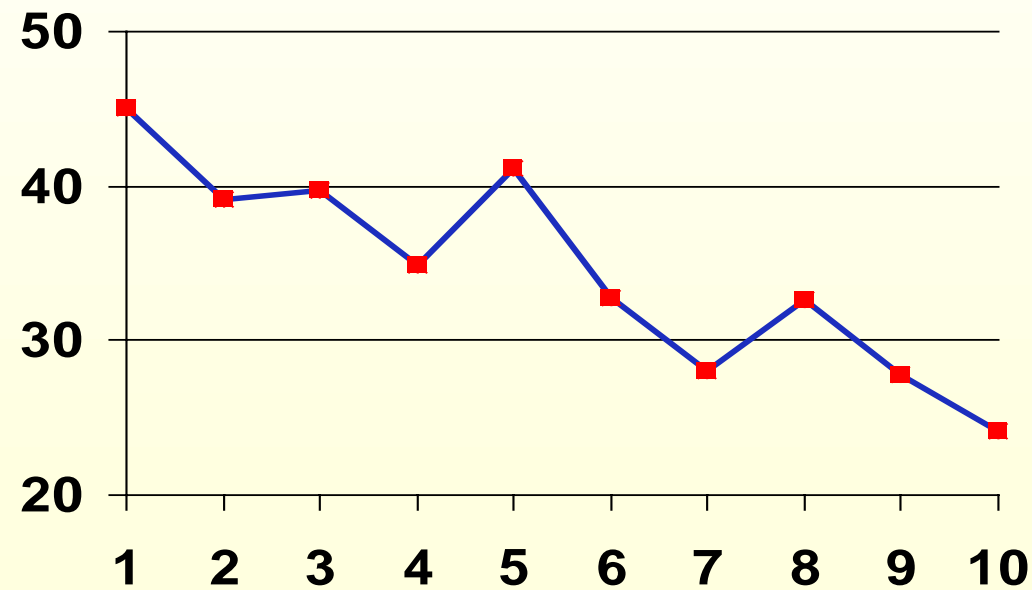
10 försök

Trial No	Initial temp Deg	Initial time min	Temp gradient Deg/min	Max temp Deg	Resolution #1 R	Resolution #2 R	Total run time min	Used Membership
1	185	5,0	5,0	295	1,201	1,959	45,063	0,417
2	195	5,0	7,0	295	1,119	1,811	39,161	0,452
3	195	5,0	5,0	305	1,226	2,02	39,69	0,527
4	185	3,0	7,0	305	1,135	1,87	34,823	0,547
5	195	3,0	5,0	295	1,182	1,992	41,1	0,488
6	200	3,0	7,0	305	1,115	1,82	32,733	0,563
7	208	2,0	8,0	310	1,043	1,725	28,009	0,596
8	197	1,5	5,5	313	1,125	1,88	32,683	0,584
9	198	2,8	7,8	322	1,08	1,735	27,786	0,613
10	199	2,6	9,1	335	1,016	1,698	24,183	0,632

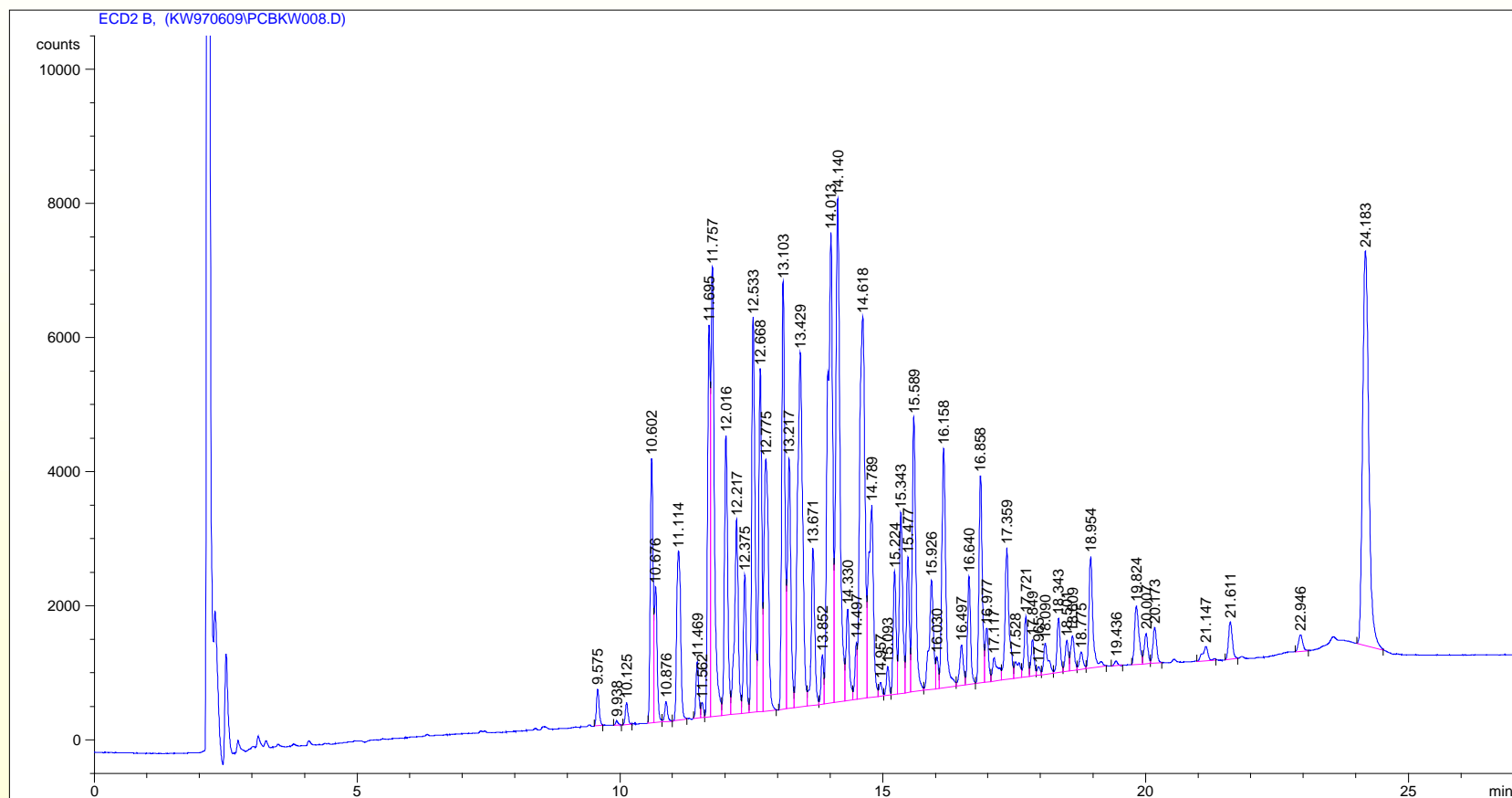
Analystiden förkortades från 45 till 25 minuter utan nämnvärd försämring av upplösningen.

Analystid vs. försök

Analystid i minuter



Kromatogram 2



Ett applikationsexempel från energiproduktion

- Miljöavgifterna på kväveoxider, 40 kr/kg, har medfört behov av åtgärder i många förbränningsanläggningar.
- Driftoptimering är ofta den mest kostnadseffektiva åtgärden att starta med.
- Simplexoptimering är en snabb och enkel metod.



Uppdraget

- Driftoptimering av ett 100 MW kraftvärmeverk, som eldar en blandning av biobränsle och olja.
- Kraftvärmeverket var utrustat med låg- NO_x -brännare med en utsläppsnivå på 85-95 mg/MJ.

Experiment- och responsvariabler

- Fyra experimentvariabler
 - Sekundärluft på två nivåer.
 - Fördelning av brännarlucht.
 - Bränslefördelning mellan brännarna.
- Två responsvariabler
 - Utsläpp av NO_x .
 - Driftsbetingelser (baserat på både mätningar och observationer).

Optimeringsförsöken

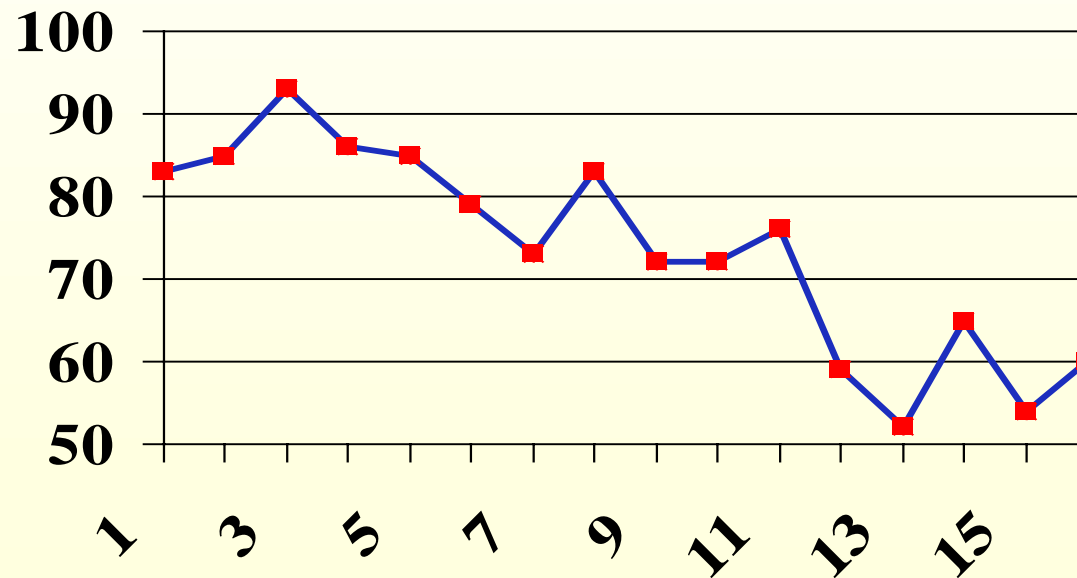
- 16 försök beräknades med den modifierade simplexalgoritmen.
- Utsläppet reducerades till 60-65 mg/MJ med bibehållet goda driftsbetingelser.
- Den 25%-iga reduktionen motsvarar en besparing på 450 SEK/h eller MSEK 2,2/år (med 5000 driftstimmar).



”Både **guld** och **gröna skogar**”

Utsläpp vs. försök

Utsläpp av NOx mg/MJ



”Evolutionär tillverkning”

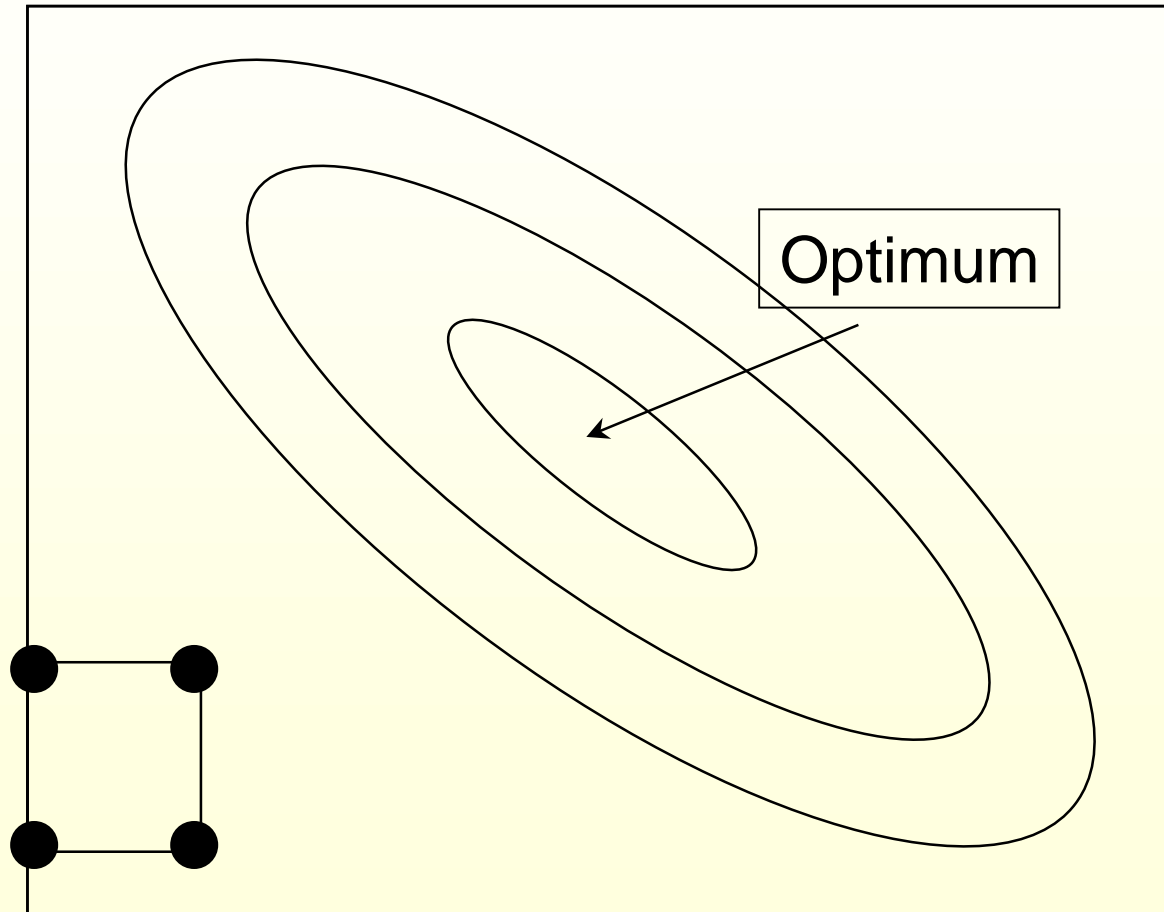
- En process är inte bara en tillverkningsanläggning. Den är också ett experimentellt system, där man kan studera effekten av olika åtgärder.
- Redan i slutet på 50-talet föreslog statistikern Georg Box, då vid ICI, att man skulle tillåta små variationer i den normala driften. Syftet är då att upptäcka det bästa sättet att driva tillverkningen.

Selektion och urval

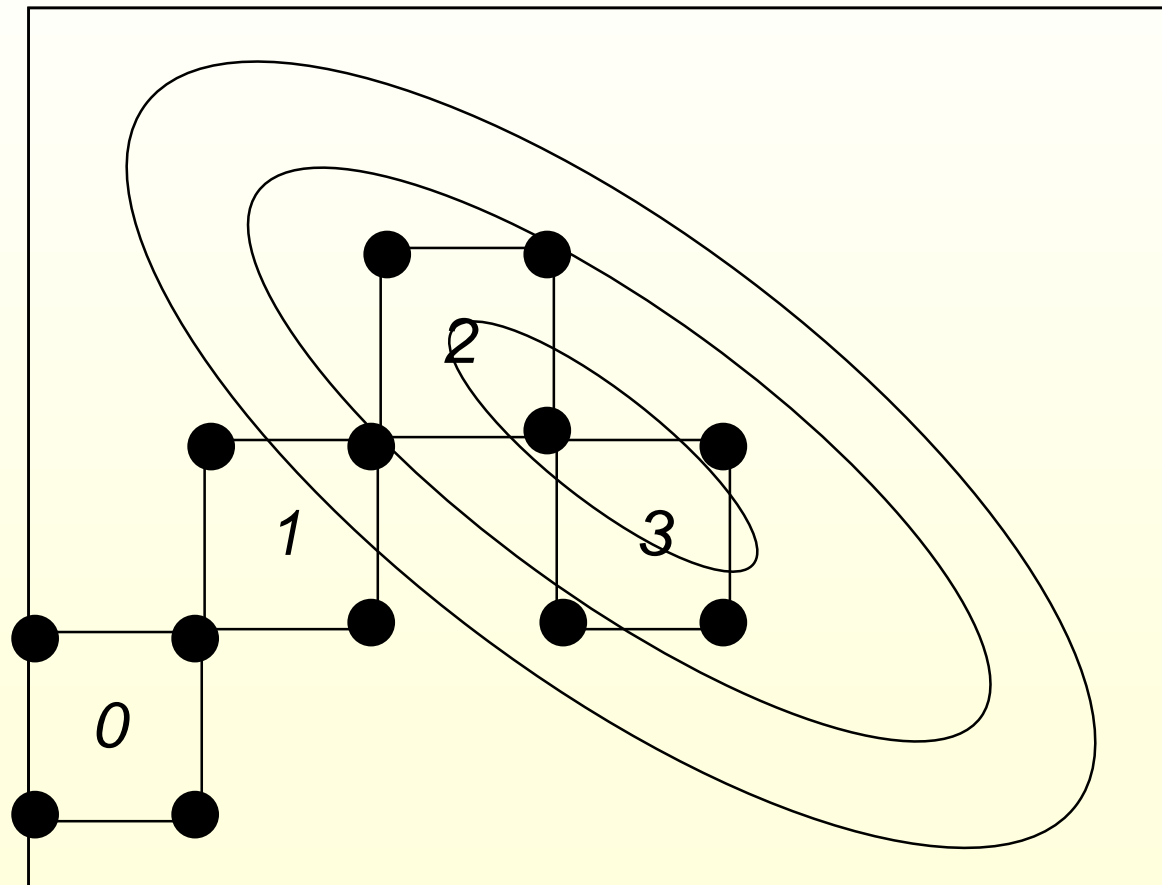
- Selektion eller urval av bättre betingelser åstadkoms genom att presentera resultaten så att det kontinuerligt går att förflytta sig mot ett optimum (enkla faktorförsök med två eller tre variabler).
- Processens naturliga variation kan skymma effekten av små variationer. I den ursprungliga utformningen upprepas därför de olika inställningarna tillräckligt många gånger för att undertrycka brusnivån.
- Signal-/brusförhållandet förbättras i proportion till:

$$\sqrt{\text{försök}}$$

Utgångsläge: Faktor-EVOP



Tre "generationer" senare

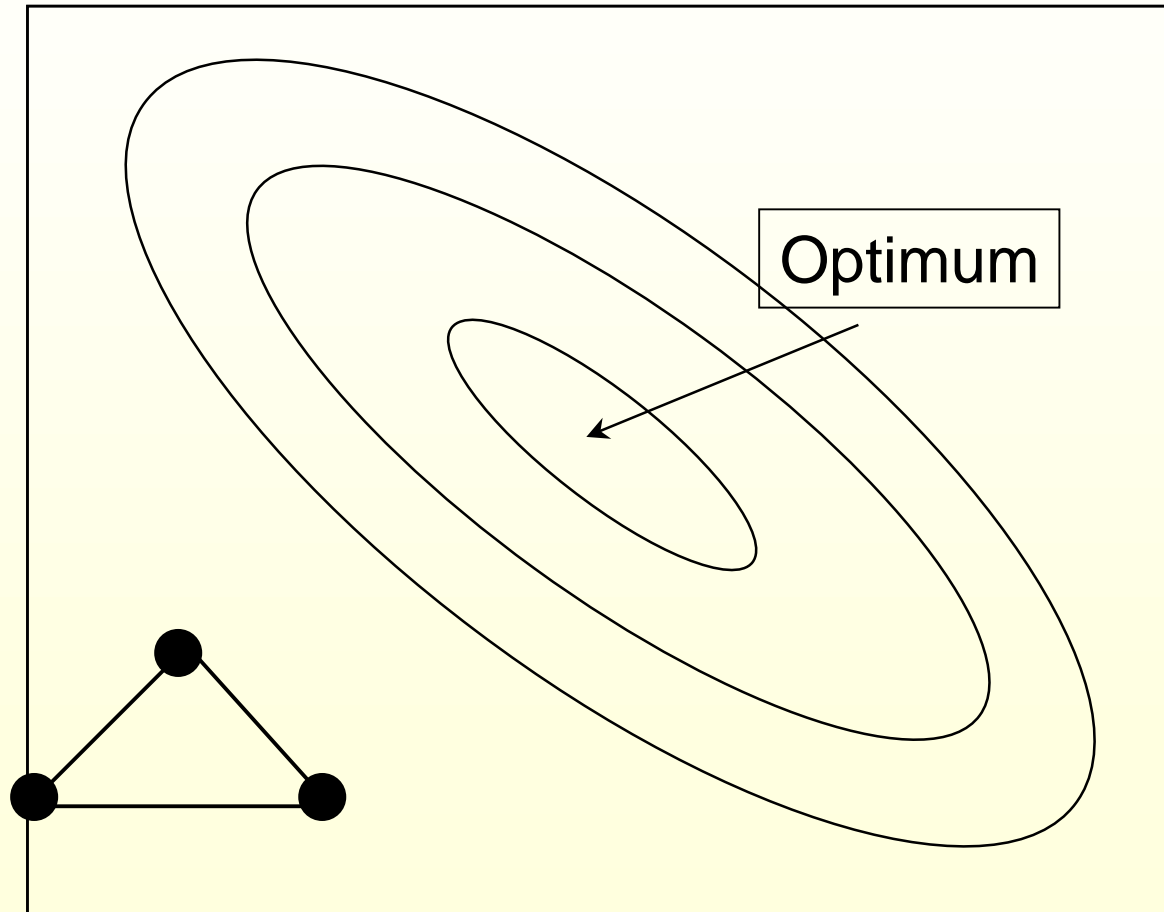


Simplex-EVOP

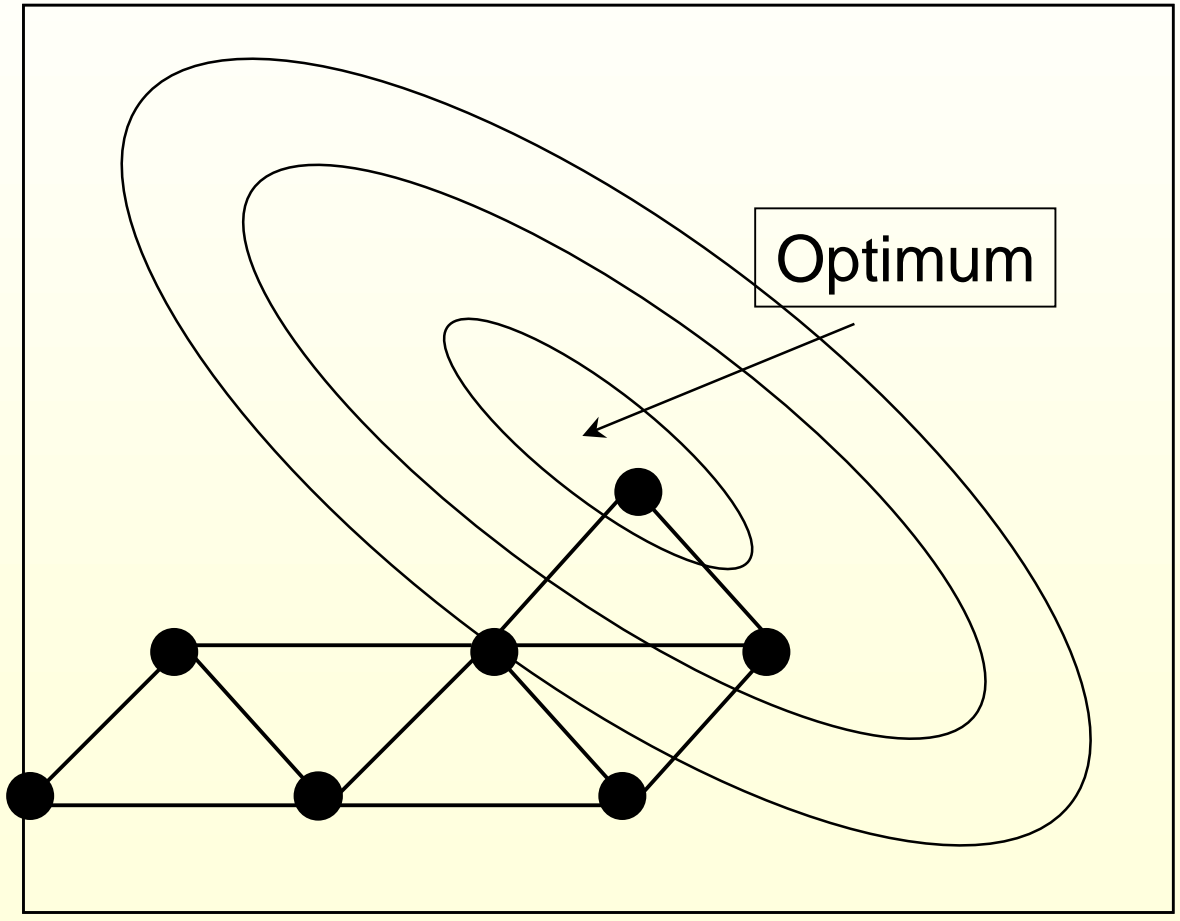
- Simplexmetoden togs fram för att automatisera EVOP-tekniken.
- Simplexmetoden är snabbare än standard-EVOP, kan hantera fler variabler och bygger ej på upprepning.
- Spendley et al visade att:
 - Prestanda i närvaro av brus ökar med antalet dimensioner.
 - Upprepningar och signifikanstestning inte förbättrar prestanda.

	Snabbhet	Antal faktorer	Automatisering	Kunskapsuppbyggnad
Faktor-EVOP	-	-	-	+
Simplex-EVOP	+	+	+	-

Utgångsläge: Simplex-EVOP



Direkt på målet: Simplex-EVOP



Automatiserad optimering

- Kontinuerlig multivariat optimering av processer och instrument kan genomföras på två principiellt olika sätt:
 - Icke-modellbaserad.
 - eller
 - Modellbaserad.
- Vilka för- respektive nackdelar erbjuder detta?



Jämförelse

- Simplex:
 - Automatisk "EVOP".
 - Störningstålig.
 - Inga modellantaganden.
 - Ingen modell som behöver uppdateras.
- Modellbaserad:
 - Känslig för störningar.
 - Antaganden om modellsamband.
 - Modell som regelbundet behöver uppdateras.

Vilken metod ska användas?

- Simplex:
 - Processer som är svåra att beskriva med enkla modeller.
 - Processer där modellsambanden förändras (t.ex. råvarukvalitet).
- Modellbaserat:
 - Processer som kan beskrivas med enkla modeller.
 - Processer där modellsambanden inte förändras.