

MODELISATION et OPTIMISATION

Claude HOINARD

**Faculté des Sciences Pharmaceutiques
« Philippe Maupas » -Tours**

Méthodologie dans l'étude d'un phénomène

Lors de l'étude d'un phénomène, plusieurs questions se posent, auxquelles répondent différents types de plans. On peut distinguer 3 grandes étapes dans l'acquisition des connaissances :

① **Recherche des facteurs influents**

② **Modélisation** : quand les facteurs influents ont été identifiés et leur importance quantifiée, on recherche ensuite l'équation permettant de décrire les variations de la réponse étudiée en fonction de celles des facteurs influents

③ **Optimisation** : déterminer quelles conditions expérimentales (les valeurs prises par les facteurs influents) permettent d'obtenir le meilleur résultat pour la réponse

⇒ la recherche des facteurs influents

Parmi tous les facteurs susceptibles d'influer sur le phénomène (c'est à dire sur la ou les réponses mesurées du phénomène),

- lesquels ont une influence significative ?
- que vaut quantitativement cette influence ?
- existe-t'il des interactions entre facteurs ?

Les plans qui permettent de rechercher les facteurs influents sont les ***plans factoriels complets ou fractionnaires***.

⇒ la modélisation

Quand les facteurs influents ont été identifiés et leur importance quantifiée, on recherche ensuite l'équation permettant de décrire les variations de la réponse étudiée en fonction de celles des facteurs influents ; cette seconde étape constitue la *modélisation*.

Les plans factoriels suffisent parfois. Il se peut dans d'autres cas que l'on soit obligé de faire appel à des plans plus complexes tels que les **plans composites centrés**.

⇒ l'optimisation

Pour finir, on cherche en général à déterminer quelles conditions expérimentales (les valeurs prises par les facteurs influents) permettent d'obtenir le meilleur résultat pour la réponse. Cette étape porte le nom d'*optimisation*.

Il existe plusieurs méthodes pour ce faire ; citons,
→ l'étude des courbes iso réponses
→ la méthode du simplexe

I - Généralités

- *Pour un système (industriel, de laboratoire...) dont l'état dépend de **variables opératoires** le terme **OPTIMISATION** désigne une **action** qui vise à trouver l'ensemble des **valeurs de ces variables** qui entraîne un **état souhaité** pour le système étudié.*
- Les problèmes d'optimisation sont très divers du point de vue des applications : optimisation d'un procédé industriel, optimisation des conditions d'emploi d'une méthode d'analyse, optimisation des coûts d'un service, optimisation de la circulation dans une ville ...etc.

Exemples

- *Dans une unité de production, **optimiser** c'est par exemple trouver les **valeurs des réglages** qui permettent d'obtenir la **meilleure qualité du produit** fabriqué*
- *Dans un laboratoire de contrôle, on peut **optimiser** l'analyse d'un produit par HPLC (réponse : surface des pics) en trouvant les **valeurs des paramètres** influents (pH, température, temps de dégazage...) qui permettent d'obtenir la **meilleure sensibilité** (plus grandes surfaces des pics)*

Dans tout problème d'optimisation, il intervient 2 catégories de variables.

- **Les variables opératoires (X) appelées aussi les facteurs.**

Ces facteurs doivent être de nature **quantitative** (*ex : pH, température...*) et il doit être possible de les faire varier à volonté (**facteurs contrôlés**) dans les limites permises par l'expérimentation.

- **La réponse (y) dite aussi fonction de réponse.**

Ce doit être aussi une variable **quantitative** ; c'est elle qui sera dite « optimisée ». La dénomination « fonction de réponse » indique que **y** doit dépendre des valeurs **X** prises par les variables opératoires.

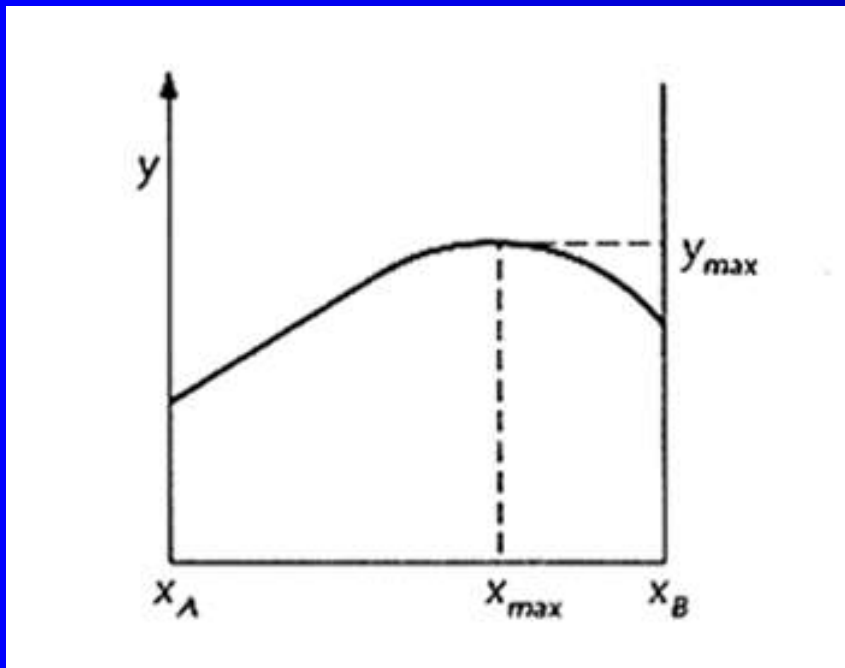
- Bien que le qualificatif « *état souhaité pour le système* » n'implique pas nécessairement la détermination d'un extremum (on peut par exemple être satisfait dans certaines études par une réponse y supérieure à une valeur seuil prédéterminée ...), nous privilégierons dans la suite du chapitre *l'optimisation* comme la *recherche d'un maximum* (ex : rendement réaction) *ou d'un minimum* (ex : coût d'une opération) de la fonction de *réponse y* en faisant varier convenablement les *variables opératoires x* (les facteurs) pour faciliter cette localisation.

- Il s'agit donc le plus souvent de trouver l'**optimum vrai** de la réponse, soit d'un point de vue mathématique de rechercher l'**extremum** (minimum ou maximum) d'une fonction y de plusieurs variables :

$$y = f(x_1, x_2, \dots, x_n)$$

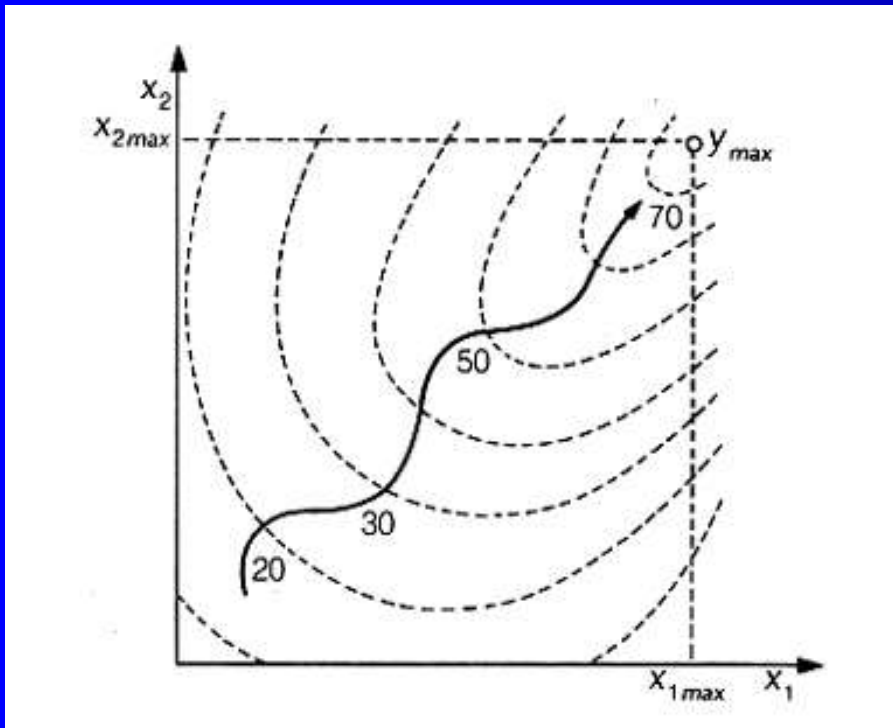
- Cas le plus simple : **une seule variable x**
- Cas le plus fréquent : **plusieurs variables opératoires, 2, 3, voire 4 ou 5**

Optimisation à une variable opératoire



- La réponse y est fonction d'une seule variable x et on recherche la valeur x_{max} de x , comprise entre les limites expérimentales x_A et x_B , qui rend la valeur de y optimale, c'est à dire y_{max} .

Optimisation à deux variables opératoires



- La fonction objectif y dépend de 2 variables x_1 et x_2 ; elle est représentée sous forme de courbes de même niveau de réponse, dites **courbes d'isoréponses** (figurées en pointillés).

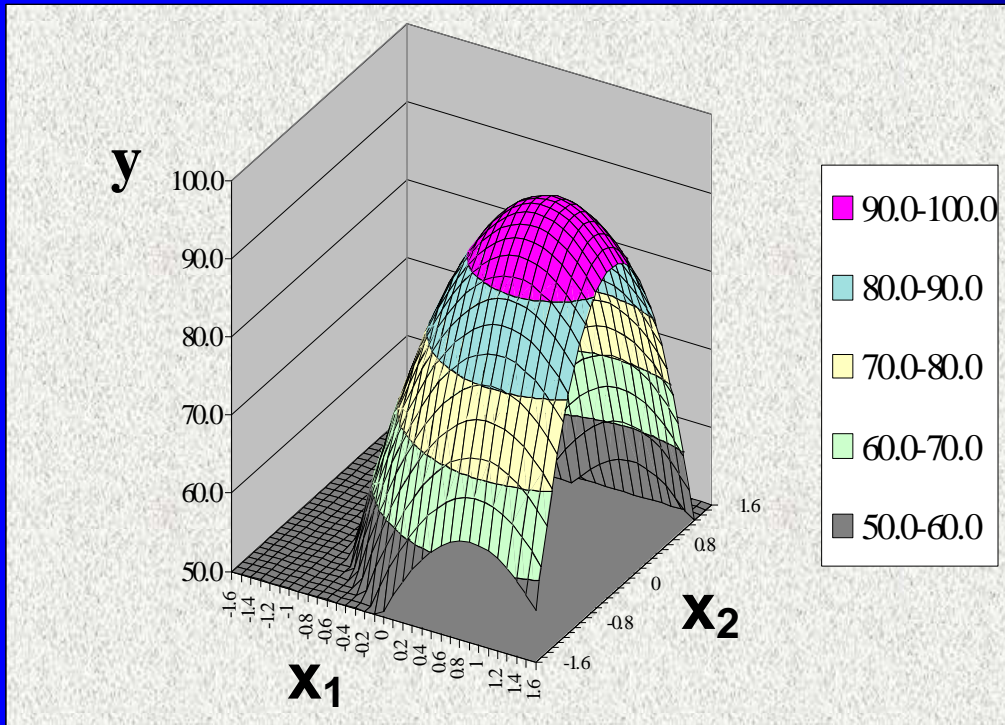
On recherche donc les valeurs x_{1max} et x_{2max} qui correspondent à la valeur y_{max} optimale.

Stratégie expérimentale d'optimisation

Quand plusieurs variables opératoires \mathbf{x} agissent sur une réponse \mathbf{y} à optimiser,

- La méthode intuitive qui consiste à faire varier « *un seul facteur à la fois en maintenant tous les autres constants* » n'est pas efficace : elle nécessite d'une part d'avantage d'essais et d'autre part on risque de ne pas atteindre l'optimum s'il existe des interactions entre les facteurs
- il est recommandé de *faire varier simultanément tous les facteurs de façon organisée*, à l'aide d'une méthodologie adéquate.

La notion de surface de réponse



Les **courbes d'isoréponses** sont les projections sur le plan $[OX_1X_2]$ de la **surface de réponse** représentant la variation de y en fonction de X_1 et X_2 dans l'espace à 3 dimensions. Notons que cette notion de surface de réponse se généralise au cas de plus de 2 facteurs.

⇒ *Le principe de toutes les méthodes d'optimisation consiste à explorer cette surface de façon à localiser un éventuel extremum dans un domaine expérimental donné.*

Classification des techniques d'optimisation

2 grandes catégories de méthodes :

- **La méthode des plans d'expériences**
méthode indirecte : modélisation puis optimisation

Déjà étudiée à propos des plans factoriels, cette méthode vise à estimer l'équation de la surface de réponse avec le modèle le plus simple possible, ***nécessairement ici au moins du 2ème degré pour restituer un maximum ou un minimum***. La position de cet optimum se déduit ensuite par dérivation de l'équation et donne les valeurs des variables opératoires.

- **La méthode du simplex et ses variantes**
méthode directe : optimisation

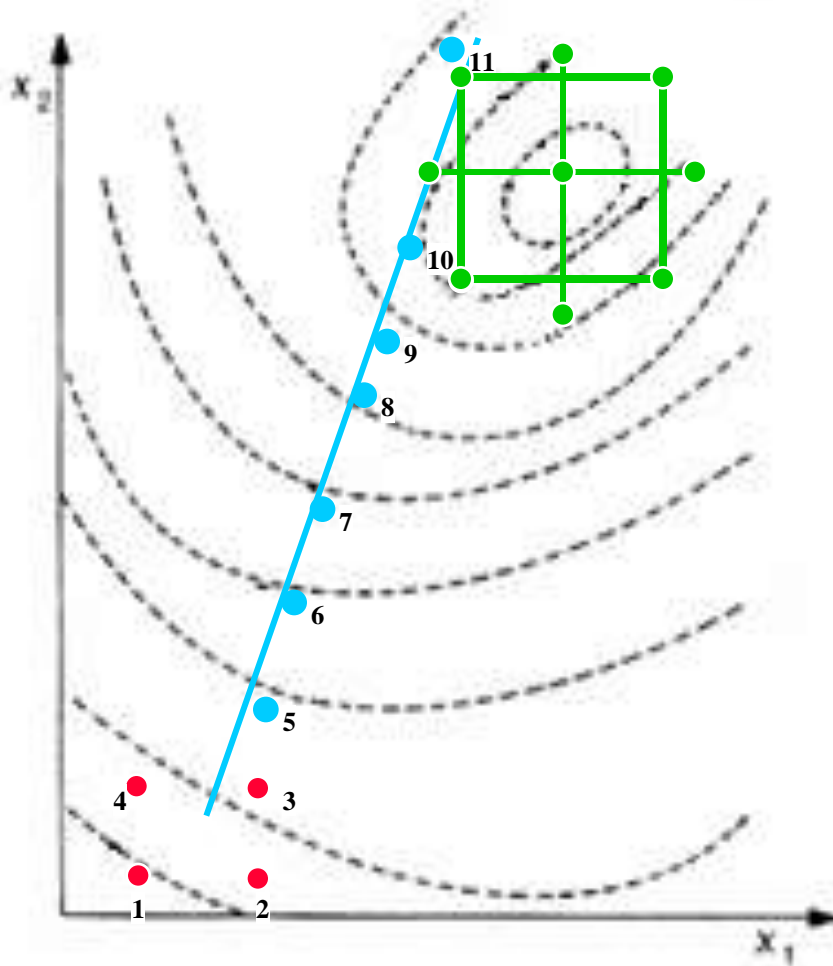
L'optimum est localisé grâce à un **algorithme itératif** : la progression vers celui-ci sur la surface de réponse se fait pas à pas, ***par étapes successives***, sans jamais avoir à déterminer l'équation de la surface.

II - La méthode des plans d'expériences

- Dans ce cas, il faut postuler ***a priori* un modèle** pour la surface de réponse puis **estimer son équation** par la technique statistique de **régression linéaire multiple**.
- La valeur des coefficients de l'équation ainsi que leur précision dépendent fortement de l'organisation des essais, d'où la **nécessité de planifier les expériences**.
- Parmi les divers plans mis au point par les statisticiens, plans de **DOEHLERT**, plans de **BOX**, **plans composites centrés (PCC)** ..., nous ne détaillerons que ces derniers parce que ce sont les plus utilisés.

Cheminement vers l'optimum

(cas avec 2 facteurs x_1 et x_2
courbes d'isoréponses en pointillé)



Localisation MAXIMUM

□ Plan Composite Centré

meilleure réponse

□ Expérimentation ligne de plus grande pente

direction optimum

équation 1er degré

facteurs/interactions avec effet

□ Plan factoriel complet (ou fractionnaire)

Quelques précisions

- **Ligne de plus grande pente.**
 - indique le chemin le plus rapide pour progresser vers l'optimum
 - équation : obtenue à partir de l'équation du 1er degré du plan factoriel initial
 - C'est une droite en l'absence d'interaction, sinon une courbe
- Dans la suite nous supposerons avoir déjà une bonne idée de la région où se situe l'optimum et réaliserons directement le **Plan Composite Centré** qui la couvre

II - 1 : les modèles mathématiques

Les modèles utilisés appartiennent tous au **modèle linéaire**, ce qui signifie que les coefficients a_i et a_{ij} de l'équation interviennent au 1er degré. Ce modèle peut être du 1er ou du 2ème degré en \mathbf{x} .

- **Le modèle du 1er degré** (en \mathbf{x}) : *des rappels.*

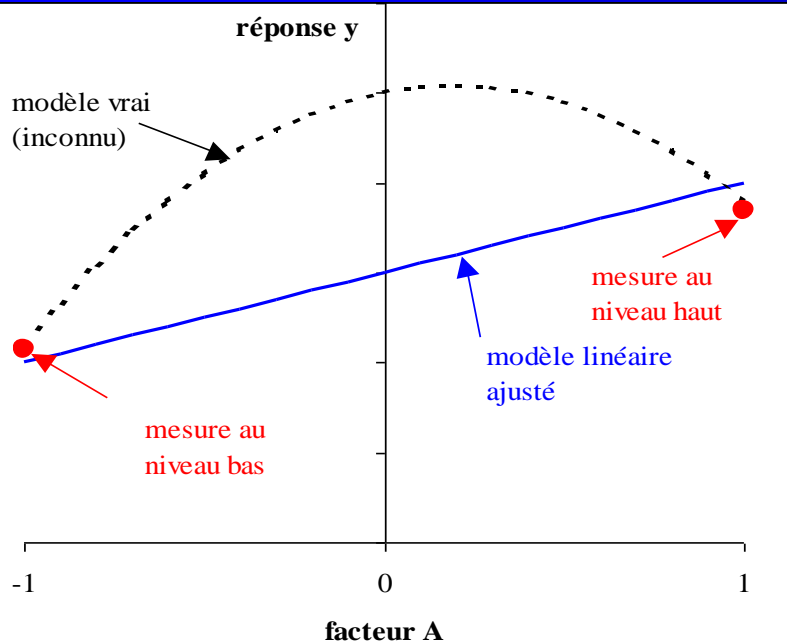
Dans le cas le plus simple de 2 variables opératoires \mathbf{x}_1 et \mathbf{x}_2 (plan factoriel 2^2), la réponse estimée \hat{y} s'écrit :

$$\hat{y} = y_c + a_1x_1 + a_2x_2 + a_{12}x_1x_2$$

a_1 et a_2 sont les effets principaux, a_{12} l'effet d'interaction et y_c la moyenne des 4 réponses, estimation de la réponse au centre du domaine expérimental.

Validité du modèle du 1er degré

□ Dans les plans factoriels 2^n chaque facteur n'est expérimenté qu'en 2 points, aux extrémités -1 et $+1$ du domaine. Cela explique le choix empirique d'une équation du 1er degré en fonction des facteurs centrés réduits X : par 2 points, il passe une droite et une seule ; mais il faut aussi ajouter qu'il peut y passer une infinité de courbes d'équations diverses ...



□ Il est donc important de s'assurer de la linéarité de l'équation dans tout le domaine en réalisant des expériences complémentaires avec un 3ème point à l'intérieur du domaine. On choisit généralement le centre ($X = 0$) situé à égale distance des extrémités expérimentées.

Validité du modèle du 1er degré

En faisant des **répétitions au centre** du domaine expérimental, on peut donc **juger la linéarité** de l'équation de prédiction en comparant statistiquement

- la moyenne y_0 de ces répétitions.
- la réponse prédite au centre par l'équation linéaire, égale à la moyenne \bar{y} des n réponses du plan factoriel 2^n .

Ces 2 moyennes doivent être en théorie égales et en pratique peu différer lorsque le modèle linéaire est valide.

Et, lorsqu'on obtient des valeurs très différentes, s'écartant de plusieurs fois l'écart type, cela signifie que le modèle linéaire n'est pas valable et qu'il faut envisager un modèle empirique plus complexe, où les facteurs centrés réduits interviennent au 2ème degré par exemple.

Notons qu'il existe des tests statistiques d'écart à la linéarité du modèle

Exemple modèle 1er degré

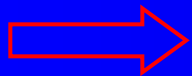
E_A , E_B et E_{AB} sont significatifs : $\hat{y} = \bar{y} + E_A X_A + E_B X_B + E_{AB} X_A X_B$

Les effets calculés ont pour valeurs

$$E_A = 6,25 \quad E_B = 11,25 \quad E_{AB} = 1,25 \quad \text{avec} \quad \bar{y} = 76,25$$

Numériquement , pour des rendements y exprimés en % , on écrit :

$$\hat{Y} = 76,25 + 6,25 X_A + 11,25 X_B + 1,25 X_A X_B$$



Cette équation permet de prédire les valeurs des rendements pour des conditions expérimentales situées à l'intérieur du domaine exploré par le plan.

Exemple : test du modèle 1er degré

Erreur-type : 0,45

moyenne calculée : $y_c = 76,25 \pm 0,45$ ($m \pm S_E$)

6 essais au centre ont été effectués pour lesquels la réponse moyenne des mesures est

moyenne mesurée : $y_o = 78 \pm 0,37$

Il y a désaccord entre la vraie moyenne mesurée et son évaluation par calcul ; un **test t** permet de le confirmer :

$$d = 78 - 76.25 = 1.75$$

$$s_d^2 = (0,37)^2 + (0,45)^2$$

$$S_d = 0,58 \quad (5+5=10 \text{ ddl})$$

$$t = d / S_d = 3,02 \quad (\text{seuil} = 2,23)$$

Le modèle du 1er degré n'est pas valide ici : toutes les prévisions de réponses sont fausses : il aurait fallu valider le modèle avant !

- **Le modèle du 2ème degré**

Le modèle du 1er degré ne peut pas convenir pour caractériser un optimum : une telle fonction ne comporte ni minimum, ni maximum.

Le modèle du 2ème degré en diffère par l'existence de termes supplémentaires du 2ème degré en \mathbf{x} :

$$\hat{y} = y_C + a_1x_1 + a_2x_2 + a_{12}x_1x_2 + a_{11}x_1^2 + a_{22}x_2^2$$

Les coefficients a_{11} et a_{22} attachés aux carrés des variables x_1 et x_2 sont primordiaux pour le calcul de l'extremum.

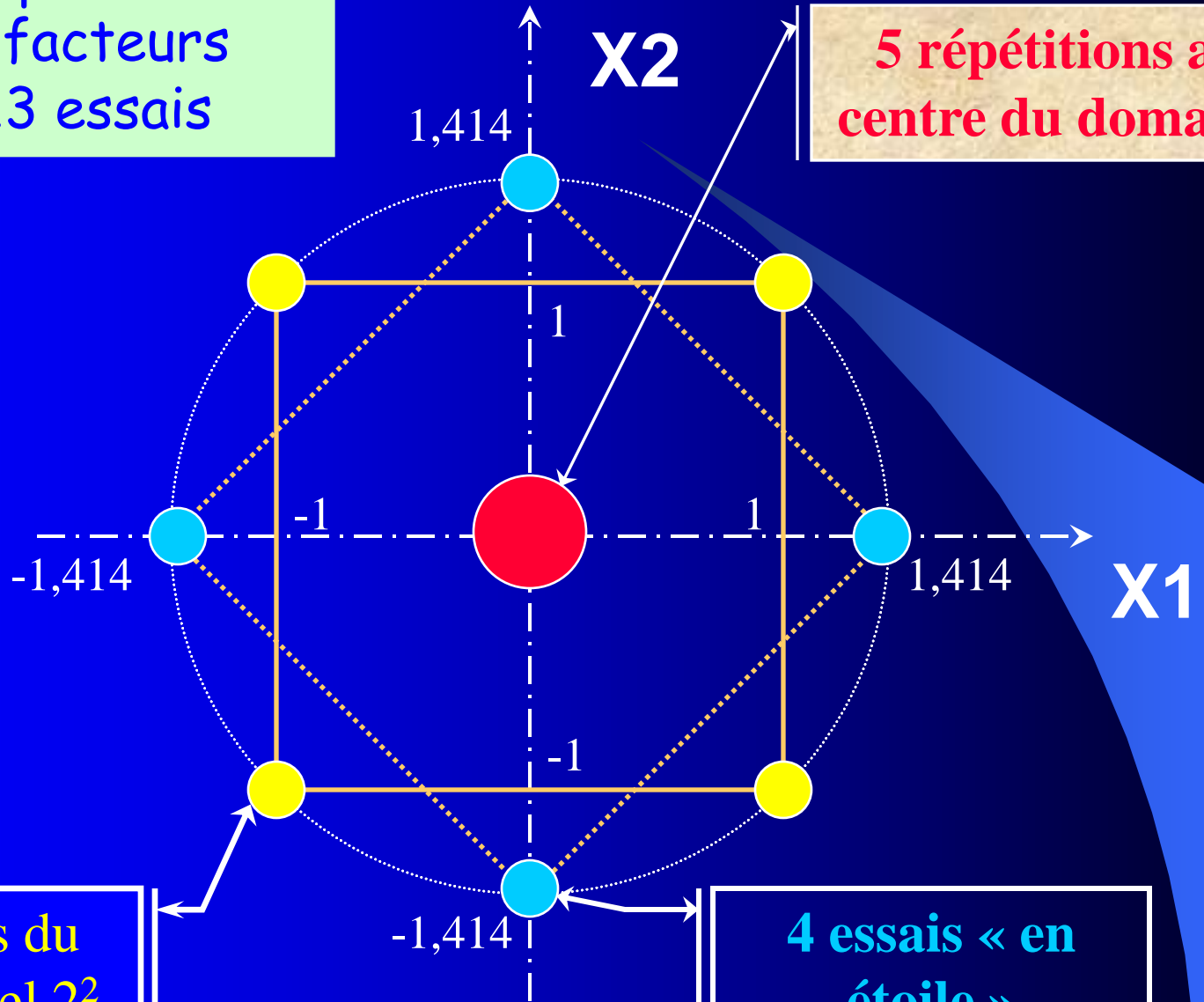
Ce modèle contient 6 paramètres (en incluant y_C) et un **minimum de 12 essais** est donc souhaitable pour appliquer convenablement la régression linéaire multiple.

*Cette condition étant remplie, il est possible de **tester la validité du modèle**, choisi empiriquement, et d'identifier les coefficients qui diffèrent statistiquement de 0 ; seuls les termes significatifs figureront dans l'équation de la surface de réponse.*

II - 2 : le plan composite centré à 2 facteurs

- Les plans composites centrés ont été conçus pour assurer une précision à peu près uniforme des estimations de réponse \hat{y}_e dans la totalité du domaine expérimenté, avec le plus petit nombre d'essais possible ...
- Ces plans comprennent 3 catégories d'essais établis de telle sorte qu'à chaque variable x correspondent 3 niveaux :
 - **les essais du plan factoriel** (pour 2^n et $n = 2$, il y en a **4**)
 - **des essais « en étoile »** par rapport aux essais précédents (au total $2n$, soit **4** pour 2 facteurs) et dont la distance à l'origine des coordonnées dépend de n (**1,414** pour $n = 2$).
 - **des essais au centre du domaine**, 3ème niveau de chaque facteur, le nombre de répétitions augmentant avec le nombre de facteurs étudiés (pour $n = 2$, il est habituel d'en prendre **5**).

Le plan composite centré à 2 facteurs comporte 13 essais



5 répétitions au centre du domaine

les 4 essais du plan factoriel 2^2

4 essais « en étoile »

La matrice d'expérience

N° essai	X1	X2
1	-1	-1
2	1	-1
3	-1	1
4	1	1
5	-1,414	0
6	1,414	0
7	0	-1,414
8	0	1,414
9	0	0
10	0	0
11	0	0
12	0	0
13	0	0

- Les niveaux des 2 facteurs sont exprimés en variables centrées réduites.
- On reconnaît les 3 catégories d'essais : **1 à 4** pour le plan factoriel, **5 à 8** pour les points en étoile et **9 à 13** pour les essais au centre.
- Il est fréquent que l'expérience soit faite en 2 phases : d'abord essais factoriels et des essais au centre puis les autres essais.

Analyse des résultats

- **Estimation de l'équation du modèle**

A l'aide de la régression linéaire multiple ; on ne tient compte que des coefficients significativement différents de 0 (risque α choisi égal à 0,05 ou plus souvent 0,10).

- **Localisation de l'extremum** : par calcul, en dérivant y par rapport à x_1 (à x_2 constant) puis par rapport à x_2 (à x_1 constant) et en recherchant la valeur du couple $[x_1, x_2]$ qui annule les 2 dérivées.

Lorsque l'extremum calculé est en dehors du domaine expérimental (x_1 et/ou $x_2 > 1,414$), il faut refaire une autre expérimentation : ***il ne faut jamais extrapoler hors du domaine.***

- **Validation** : la prédiction de l'optimum est mathématique, à l'aide d'un modèle empirique... Il est indispensable de faire l'essai à cet optimum supposé : la réponse mesurée doit correspondre à celle qui est prédite.

Visualisation des résultats

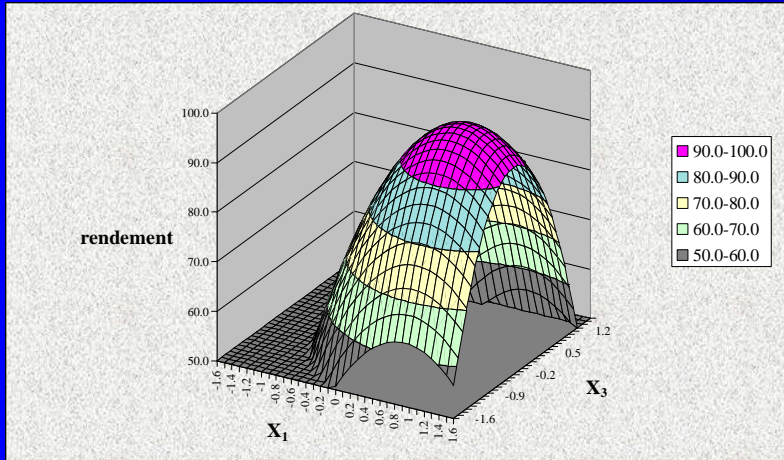


Diagramme de surface de réponse
On voit que la réponse passe par un maximum, mais ce type de graphe est difficile à exploiter

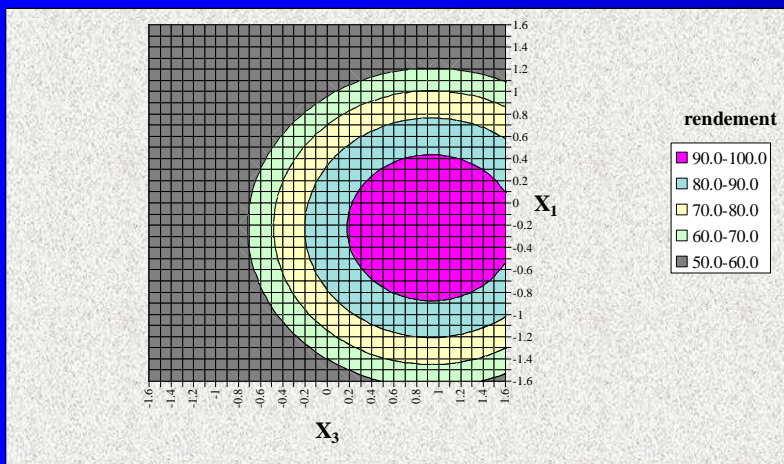


Diagramme des courbes d'isoréponses

Projection du graphe précédent (vue du dessus)

Commode pour visualiser la région du maximum avec les courbes de niveau

Exemple d 'application (exercice 1)

- Il s 'agit de l 'étude de la stabilité d 'une émulsion de composition donnée, fabriquée avec un broyeur colloïdal. Les 2 facteurs x étudiés sont :
 - la largeur de l 'entrefer (centre 1,25 mm)
 - la vitesse du rotor (centre 750 t.min⁻¹)
- L 'objectif de l 'expérimentation est d 'obtenir une réponse y (la stabilité, mesurée en points d 'indice) optimisée, qui ne se définit pas ici comme l 'estimation d 'un extremum mais se caractérise par la recherche du domaine expérimental tel que cette stabilité soit au moins égale à 90 points.

1ère phase : plan factoriel 2^2

La correspondance entre les valeurs centrées réduites des niveaux des facteurs et les valeurs réelles expérimentées sont les suivantes :

- Largeur entrefer (x_1)

x réduit	valeur
-1,414	= 0,50 mm
-1	= 0,71 mm
0	= 1,25 mm
+1	= 1,79 mm
+1,414	= 2,00 mm

- Vitesse du rotor (x_2)

x réduit	valeur
-1,414	= 600 t/min
-1	= 643 t/min
0	= 750 t/min
+1	= 857 t/min
+1,414	= 900 t/min

- Les réponses des essais 1 à 4 permettent de calculer les coefficients (les effets) du plan factoriel :

$$a_1 = -8,25 ; a_2 = 7,25 ; a_{12} = -1,25 \text{ et } y_C = 85,25$$

L'erreur type d'un coefficient est $\sigma_E = 1,5$ (puisque l'on sait que $\sigma_y = 3$). Un coefficient est significatif si sa valeur absolue dépasse $2\sigma_E$, soit 3. L'interaction n'est donc pas significative. L'équation s'écrit :

$$\hat{y} = 85,25 - 8,25x_1 + 7,25x_2$$

- Les 3 répétitions au centre (5 à 7) permettent d'obtenir une mesure directe de la réponse au centre du domaine par sa moyenne : $y_0 = 72,67$.

La différence entre cette moyenne et l'estimation y_C donnée par le modèle (12,58) est très grande par rapport à la dispersion des mesures : *le modèle du 1er degré ne convient pas.*

2ème phase : plan composite centré

- L'expérimentation a ensuite été complétée par 4 essais en étoile et 2 nouvelles répétitions au centre ; on dispose alors des 13 réponses du plan CC. On ajuste ensuite une équation du 2ème degré comprenant 6 coefficients (cf 12) au moyen de la régression linéaire multiple. Les résultats sont les suivants :

coefficient :		erreur type	t de student	degré de signification
nature	valeur			
y_c	72,80	1,17	62,15	$7 \cdot 10^{-11}$
a_1	-9,25	0,93	9,99	$2 \cdot 10^{-5}$
a_2	6,63	0,93	7,16	$2 \cdot 10^{-4}$
a_{11}	7,29	0,99	7,34	$2 \cdot 10^{-4}$
a_{22}	6,29	0,99	6,33	$4 \cdot 10^{-4}$
a_{12}	-1,25	1,31	0,954	0,372

Seule l'interaction a_{12} n'est pas significative

- **L'estimation de l'équation de la surface de réponse** a pour expression :

$$\hat{y} = 72,80 - 9,25x_1 + 6,63x_2 + 7,29x_1^2 + 6,29x_2^2$$

Sa représentation graphique est une surface parabololoïde qui présente une concavité dirigée vers le bas (la fonction présente donc un minimum).

- **Validité de cette équation.** Elle est examinée de 2 façons :
 - par l'étude des résidus individuels bruts (écart entre valeur mesurée et valeur estimée) ou mieux par les résidus normalisés ; ces derniers doivent en principe être inférieurs à 2. Supérieur à 2,5, un résultat devient suspect. Dans l'exemple étudié, aucun résidu n'est suspect.
 - par un test global d'adéquation du modèle utilisé aux données mesurées, inclus dans l'analyse de variance de la régression. Ce test est effectué par *Minitab* (lack of fit). Dans l'exemple, il est loin d'être significatif (P = 0,47).

- **Utilisation de l'équation**

Elle permet le calcul de la stabilité prévisible \hat{y} pour des valeurs données des 2 variables opératoires.

Pour $x_1 = 0,8$ mm

$x_2 = 800$ t /min

$$x_1 = (0,8 - 1,25) / 0,54 = \mathbf{-0,83}$$

$$x_2 = (800 - 750) / 107 = \mathbf{+0,47}$$

Dans l'équation précédente, en remplaçant x_1 et x_2 par leur valeur, on obtient :

$$\hat{y} \# \mathbf{90}$$

Il ne s'agit certes que d'une prévision et non d'une mesure. Mais on peut considérer que celle-ci est fiable dans la mesure où le point est à l'intérieur du domaine expérimental et où l'équation du modèle est valide.

Détermination du domaine $[x_1, x_2]$ tel que $\hat{y} > 90$

- **Par un tableau à double entrée** (une grille) comportant en lignes et colonnes marginales des valeurs régulièrement espacées de x_1 et x_2 (les mailles) et à l'intérieur du tableau les valeurs des stabilités prédites \hat{y} obtenues par calcul au moyen de l'équation. Ce tableau peut servir à présenter les couples possibles pour x_1 et x_2



- **Par un graphique 3D** représentant les données du tableau, soit la **surface de réponse** elle même mais ce graphe est difficile à exploiter, soit mieux sa projection sur le plan $[OX_1, OX_2]$ qui donne **le réseau de courbes isoréponses**. Celui-ci fait apparaître le domaine pour lequel la condition de stabilité $\hat{y} > 90$ est satisfaite, ainsi que le montre le graphe suivant.

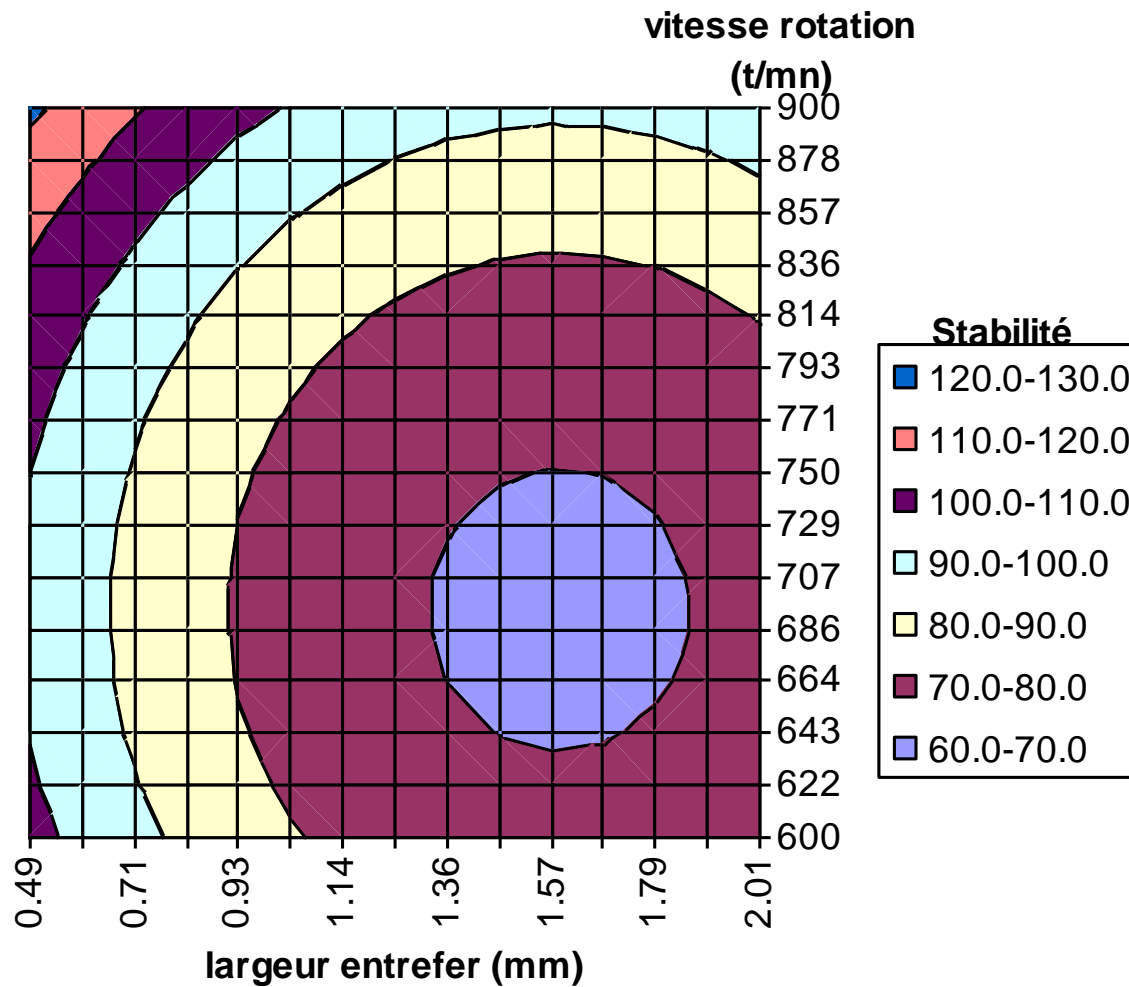


Tableau permettant de prédire les conditions de vitesse de rotation et de largeur d'entrefer donnant une stabilité > 90

		points stabilité supérieurs à 90														
		Vitesse de Rotation (t/mn)														
largeur entrefer (mm)		-1.4	-1.2	-1	-0.8	-0.6	-0.4	-0.2	0	0.2	0.4	0.6	0.8	1	1.2	1.4
		600	622	643	664	686	707	729	750	771	793	814	836	857	878	900
-1.4	0.49	103.1	101.1	99.7	98.8	98.3	98.4	99.0	100.0	101.6	103.7	106.3	109.4	113.0	117.0	121.6
-1.2	0.60	97.4	95.5	94.1	93.1	92.7	92.7	93.3	94.4	96.0	98.1	100.6	103.7	107.3	111.4	116.0
-1	0.71	92.4	90.4	89.0	88.1	87.6	87.7	88.3	89.3	90.9	93.0	95.6	98.7	102.3	106.3	110.9
-0.8	0.82	87.9	86.0	84.5	83.6	83.2	83.2	83.8	84.9	86.4	88.5	91.1	94.2	97.8	101.9	106.5
-0.6	0.93	84.0	82.1	80.6	79.7	79.3	79.3	79.9	81.0	82.6	84.6	87.2	90.3	93.9	98.0	102.6
-0.4	1.03	80.7	78.8	77.3	76.4	76.0	76.0	76.6	77.7	79.2	81.3	83.9	87.0	90.6	94.7	99.3
-0.2	1.14	78.0	76.0	74.6	73.7	73.2	73.3	73.9	74.9	76.5	78.6	81.2	84.3	87.9	92.0	96.5
0	1.25	75.8	73.9	72.5	71.5	71.1	71.2	71.7	72.8	74.4	76.5	79.0	82.1	85.7	89.8	94.4
0.2	1.36	74.3	72.3	70.9	70.0	69.5	69.6	70.2	71.2	72.8	74.9	77.5	80.6	84.2	88.3	92.8
0.4	1.47	73.3	71.4	69.9	69.0	68.6	68.6	69.2	70.3	71.8	73.9	76.5	79.6	83.2	87.3	91.9
0.6	1.57	72.9	71.0	69.5	68.6	68.2	68.2	68.8	69.9	71.5	73.5	76.1	79.2	82.8	86.9	91.5
0.8	1.68	73.1	71.2	69.7	68.8	68.3	68.4	69.0	70.1	71.6	73.7	76.3	79.4	83.0	87.1	91.7
1	1.79	73.9	71.9	70.5	69.6	69.1	69.2	69.8	70.8	72.4	74.5	77.1	80.2	83.8	87.8	92.4
1.2	1.90	75.2	73.3	71.8	70.9	70.5	70.5	71.1	72.2	73.8	75.9	78.4	81.5	85.1	89.2	93.8
1.4	2.01	77.2	75.2	73.8	72.9	72.4	72.5	73.1	74.1	75.7	77.8	80.4	83.5	87.0	91.1	95.7



Courbes d'isoréponses

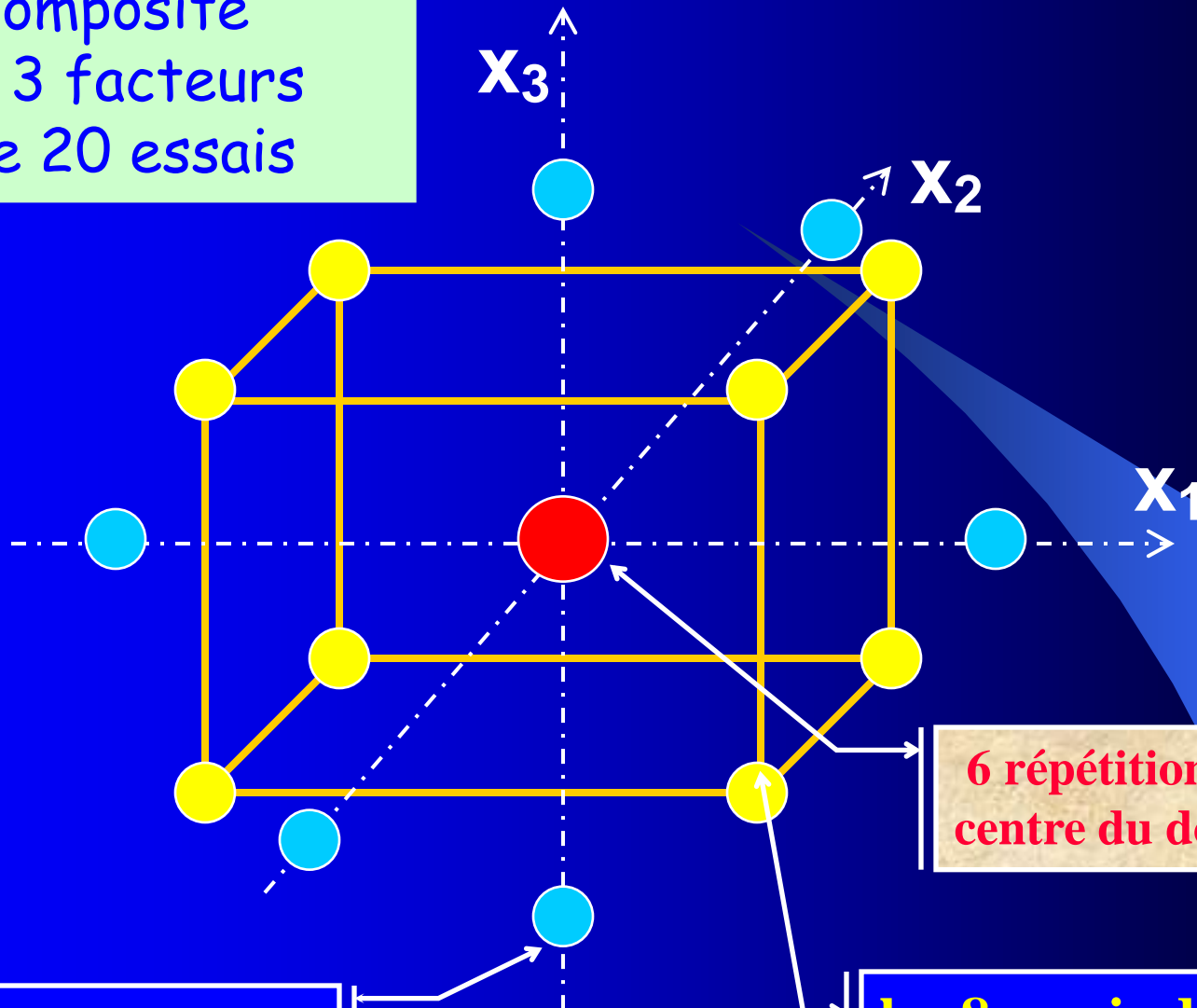


II - 3 : le plan composite centré à 3 facteurs

- Ce plan étudie l'influence de 3 facteurs (x_1 , x_2 , x_3) sur la réponse y à optimiser.
- Il comporte 3 catégories d'essais :
 - ***les essais du plan factoriel*** : 8 aux sommets du cube de côté 1.
 - ***Les essais « en étoile »*** : 6 points situés sur les 3 axes, 1 par face du cube, situés tous à la même distance de l'origine, égale à $\alpha = 1,682$ en variable réduite.
 - ***Les essais au centre*** : il faut prévoir 6 répétitions pour que la précision des estimations soit à peu près constante dans le domaine expérimental.

Au total 20 essais

Le plan composite centré à 3 facteurs comporte 20 essais



6 répétitions au centre du domaine

6 essais « en étoile »

les 8 essais du plan factoriel 2^3

L'équation du modèle

- Elle s'écrit :

$$\hat{y} = y_c + a_1x_1 + a_2x_2 + a_3x_3 + a_{12}x_1x_2 + a_{13}x_1x_3 + a_{23}x_2x_3 + a_{11}x_1^2 + a_{22}x_2^2 + a_{33}x_3^2$$

- Noter qu'on tient compte des 3 interactions d'ordre 2 (a_{12} , a_{13} et a_{23}) mais pas de l'interaction d'ordre 3. Au total, **ce modèle comporte 10 paramètres à estimer**. Comme l'expérimentation comportera 20 réponses, la précision des estimations sera convenable ($n \geq 2p$).
- Dans l'équation numérique d'une expérimentation, on ne tient compte que des termes dont le coefficient est significativement différent de 0.

Analyse et interprétation

- *Quand les 20 réponses y sont connues, il faut :*

1 - calculer les coefficients a_{ij} du modèle au moyen de la régression linéaire multiple. Ne retenir dans l'équation que les termes dont le coefficient est significativement différent de 0.

2 - Déterminer par calcul les valeurs des facteurs réduits qui rendent la réponse y optimale (dérivation) à l'aide de l'équation précédemment déterminée. Transformer ces valeurs réduites en valeurs réelles.

3 - Effectuer les représentations graphiques, surtout les courbes de niveaux en prenant les variables 2 par 2.

4 - Faire un ou plusieurs essais de confirmation à l'extremum pour valider les conclusions.

II - 4 : exemple d 'application (exercice 2)

- Il s 'agit de déterminer les conditions expérimentales d 'une réaction chimique, pour la préparation d 'un principe actif médicamenteux, qui assurent le rendement de réaction le plus élevé possible.
- 3 facteurs sont étudiés :
 - x_1 est le rapport molaire de triéthylamine $\text{Et}_3\text{N} / M_1$ où M_1 désigne la matière première de départ.
 - x_2 est la température θ_1 à laquelle est introduit un réactif M_2 agissant sur M_1 .
 - x_3 est le rapport molaire de M_2 / M_1 .

La recherche du maximum de rendement y est effectué au moyen d 'un plan composite centré, soit 20 essais.

Matrice d'expériences

Le tableau suivant permet de déterminer les conditions expérimentales à réaliser :

facteur	Valeur au centre du domaine	Pas de variation
x_1	1	0,5
x_2	15°C	8°C
x_3	1	0,5

Les valeurs centrées réduites X_1 , X_2 et X_3 qui interviennent dans le plan sont reliées à ces valeurs par :

$$X_1 = \frac{x_1 - 1}{0,5} \quad X_2 = \frac{x_2 - 15}{8} \quad X_3 = \frac{x_3 - 1}{0,5}$$

n° essai	X_1	X_2	X_3	Rendement mesuré y
1	-1	-1	-1	23
2	1	-1	-1	31
3	-1	1	-1	25
4	1	1	-1	7
5	-1	-1	1	67
6	1	-1	1	85
7	-1	1	1	69
8	1	1	1	63
9	-1,682	0	0	71
10	1,682	0	0	3
11	0	-1,682	0	75
12	0	1,682	0	87
13	0	0	-1,682	3
14	0	0	1,682	97
15	0	0	0	85
16	0	0	0	89
17	0	0	0	83
18	0	0	0	85
19	0	0	0	83
20	0	0	0	83

Les 8 essais du plan factoriel

Les 6 essais en étoile

Les 6 répétitions au centre

Résultats de la régression linéaire multiple

désignation	valeur coefficient	erreur-type	statistique t	degré de signification
constante	84.925	5.64	15.06	3.37×10^{-08}
a ₁	-8.228	3.74	-2.20	0.053
a ₂	-1.598	3.74	-0.43	0.678
a ₃	26.074	3.74	6.97	3.86×10^{-05}
a ₁₂	-6.250	4.89	-1.28	0.230
a ₁₃	2.750	4.89	0.56	0.586
a ₂₃	0.250	4.89	0.05	0.960
a ₁₁	-18.543	3.64	-5.09	4.71×10^{-04}
a ₂₂	-2.987	3.64	-0.82	0.431
a ₃₃	-13.947	3.64	-3.83	0.003

En rouge, les coefficients significatifs au risque $\alpha = 0,10$.

L 'équation ne dépend que de X₁ et X₃ et s 'écrit :

$$y = 84,925 - 8,23 * X_1 + 26,07 * X_3 - 18,54 * X_1^2 - 13,95 * X_3^2$$

$$y = 84,925 - 8,23 * X_1 + 26,07 * X_3 - 18,54 * X_1^2 - 13,95 * X_3^2$$

Détermination de l'optimum

⇒ La dérivée de y par rapport à X_1 a pour expression :

$$dy/dX_1 = -8,23 - 2 * 18,54 * X_1$$

elle s'annule pour $X_{1max} = -8,23 / 37,08 = -0,22$

⇒ La dérivée de y par rapport à X_3 a pour expression :

$$dy/dX_3 = 26,07 - 2 * 13,95 * X_3$$

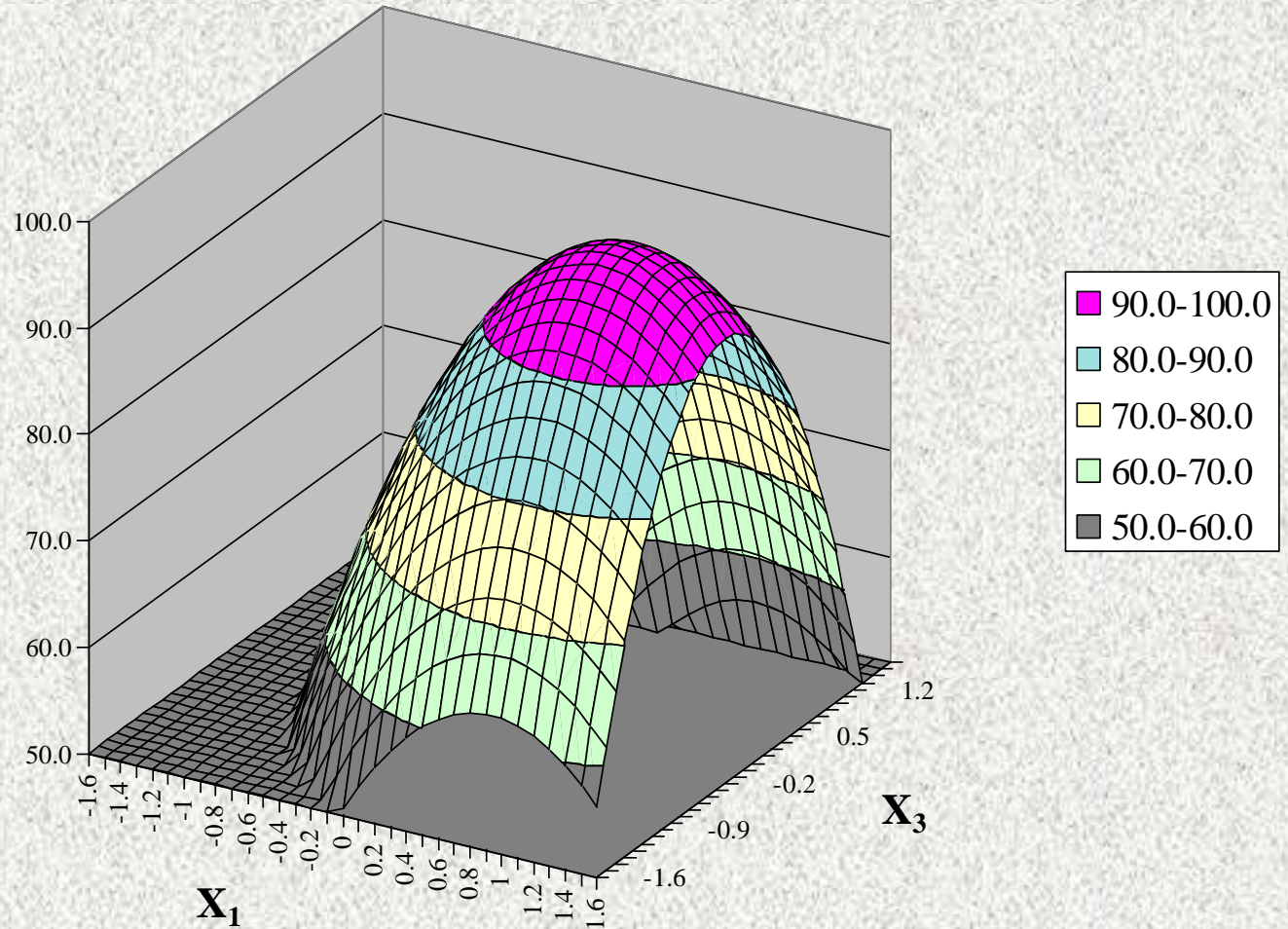
elle s'annule pour $X_{3max} = 26,07 / 27,90 = 0,93$

La valeur y_{max} prévisible pour ces valeurs optimales de X_1 et X_3 est calculée par l'équation :

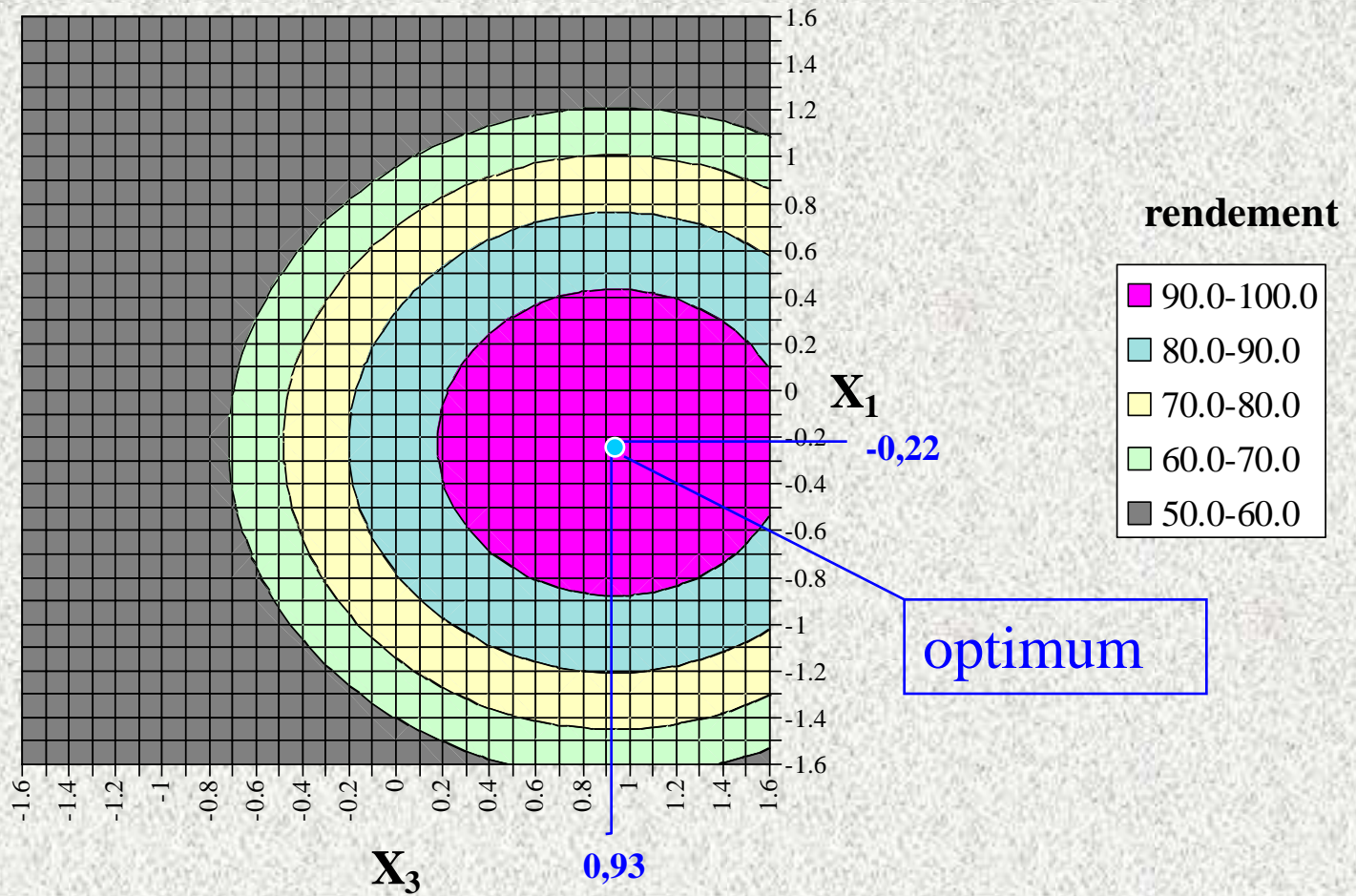
$$y_{max} = 98,0$$

Surface de réponse

rendement



Courbes d'iso-réponses



$$X_1 = \frac{x_1 - 1}{0,5} \quad X_3 = \frac{x_3 - 1}{0,5}$$

$$X_{1\max} = -0.22$$

$$X_{3\max} = 0.93$$

Une remarque pour terminer

- Les conditions opératoires réelles qui maximisent le rendement sont :

$x_1 = 0,89$ rapport molaire $\text{Et}_3\text{N} / \text{M}_1$

$x_3 = 1,47$ rapport molaire M_2 / M_1

la température x_2 n'a pas d'influence

Le rendement estimé est $y_{\max} = 98,0$

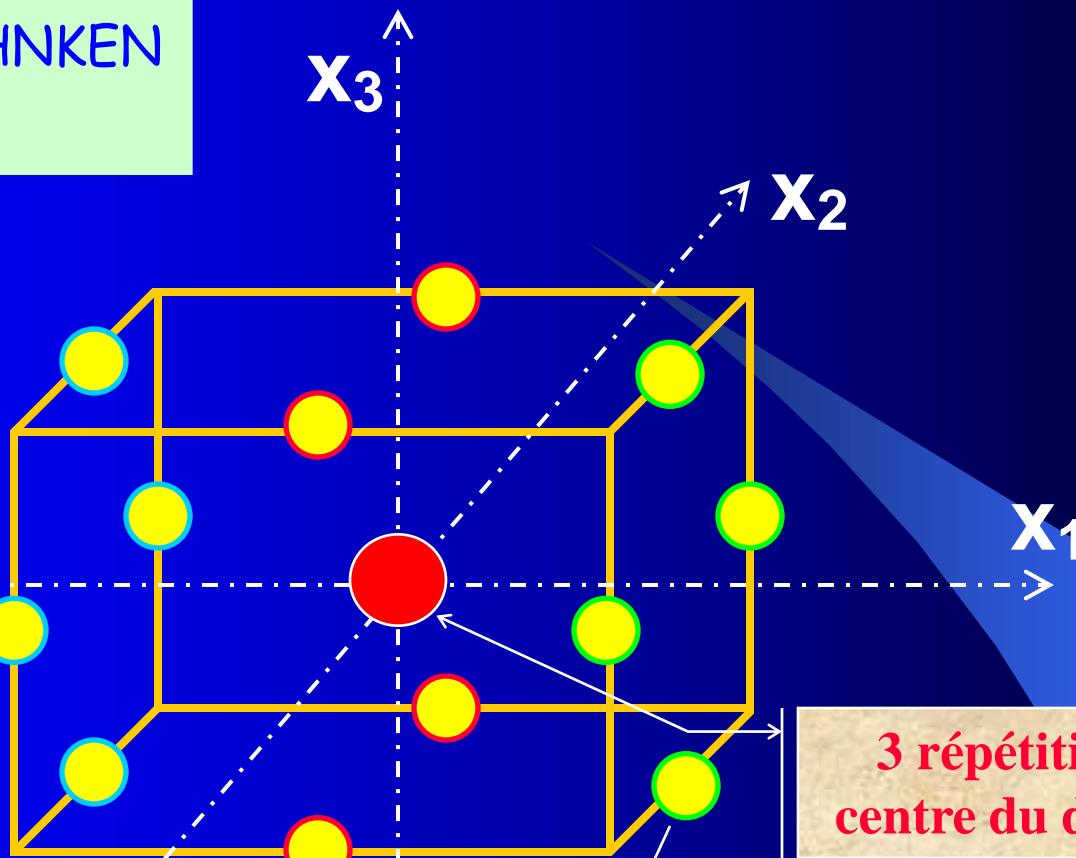
Il est absolument nécessaire de faire l'expérience dans ces conditions afin de vérifier que la mesure expérimentale du rendement est en accord avec ce qui a été prédit ...

II - 5 : aperçu sur les plans de Box-Behnken

- Il existe d'autres plans possibles pour localiser un optimum et qui correspondent à des essais différents de ceux des plans composites centrés.
- Les plans de Box-Behnken, qui figurent dans le logiciel *Minitab*, sont des plans économiques comportant moins d'essais que le plan composite centré correspondant au même nombre de facteurs pour localiser un optimum ; par exemple pour 3 facteurs, il n'y a que 15 essais à réaliser au lieu de 20. Il s'analyse par la régression linéaire multiple comme le PCC. Du point de vue statistique, ces plans sont moins sûrs et couvrent un domaine expérimental moins étendu que le plan composite ainsi que le montre le schéma ci-après :

plan de BOX-BEHNKEN
à 3 facteurs

12 points sont
répartis à la
surface du cube :



4 sur la face
 $X_1 = -1$

3 répétitions au
centre du domaine

4 sur la
section
 $X_1 = 0$

4 sur la face
 $X_1 = +1$

Soit,
au total 15 essais

III - La méthode du simplex et ses variantes

- Inventée par SPENDLEY en 1962, cette méthode consiste à **se déplacer pas à pas sur la surface de réponse** grâce à un algorithme itératif, **sans** jamais avoir à **déterminer l'équation** liant la réponse y aux variables opératoires x .
- *Une comparaison imagée* : la progression vers l'optimum se fait un peu comme celle d'un alpiniste perdu dans le brouillard, qui disposerait d'un altimètre pour savoir s'il monte ou descend (traduire la réponse augmente ou diminue) et d'une boussole qui lui indique la direction ... muni de ces 2 indicateurs, il essaie de trouver le sommet le plus rapidement possible.

III - 1 : description succincte de la méthode

- La méthode repose sur la variation simultanée des k variables opératoires X de l'expérimentation.

Un simplex est un ensemble de points déterminés de façon à former un réseau : figure géométrique à $k + 1$ sommets équidistants d'un point central dans un espace à k dimensions. Dans le cas de 2 variables le simplex est un triangle équilatéral ; dans le cas de 3 variables, le simplex est un tétraèdre régulier ...

- Le démarrage de la méthode consiste à effectuer un premier simplex, choisi au hasard, soit $k + 1$ essais qui permettent de récolter autant de réponses y ; elles sont alors triées par ordre croissant.

- La progression vers l'optimum se fait essai après essai en supprimant de l'expérimentation effectuée la réponse la plus mauvaise et en la remplaçant par un autre point situé à l'opposé de celui-ci sur l'axe passant par le centre de gravité de autres points de façon à obtenir un nouveau simplex. Ce nouveau point est expérimenté et les k réponses y comparées comme précédemment pour voir quelle est la nouvelle plus mauvaise réponse à éliminer et quelle nouvelle direction prendre. Le procédé est répété jusqu'à l'obtention d'une réponse satisfaisante.

- La régularité du simplex permet de ne privilégier aucune direction. La méthode permet des changements successifs de direction. La logique de la démarche est qu'en se déplaçant dans la direction opposée au plus mauvais résultat, on va améliorer la réponse et finalement atteindre l'optimum.

III - 2 : le simplexe à 2 variables

L'étude porte sur 2 facteurs quantitatifs x_1 et x_2 agissant sur une réponse y . Comme dans les plans expérimentaux, ces variables opératoires doivent être réduites.

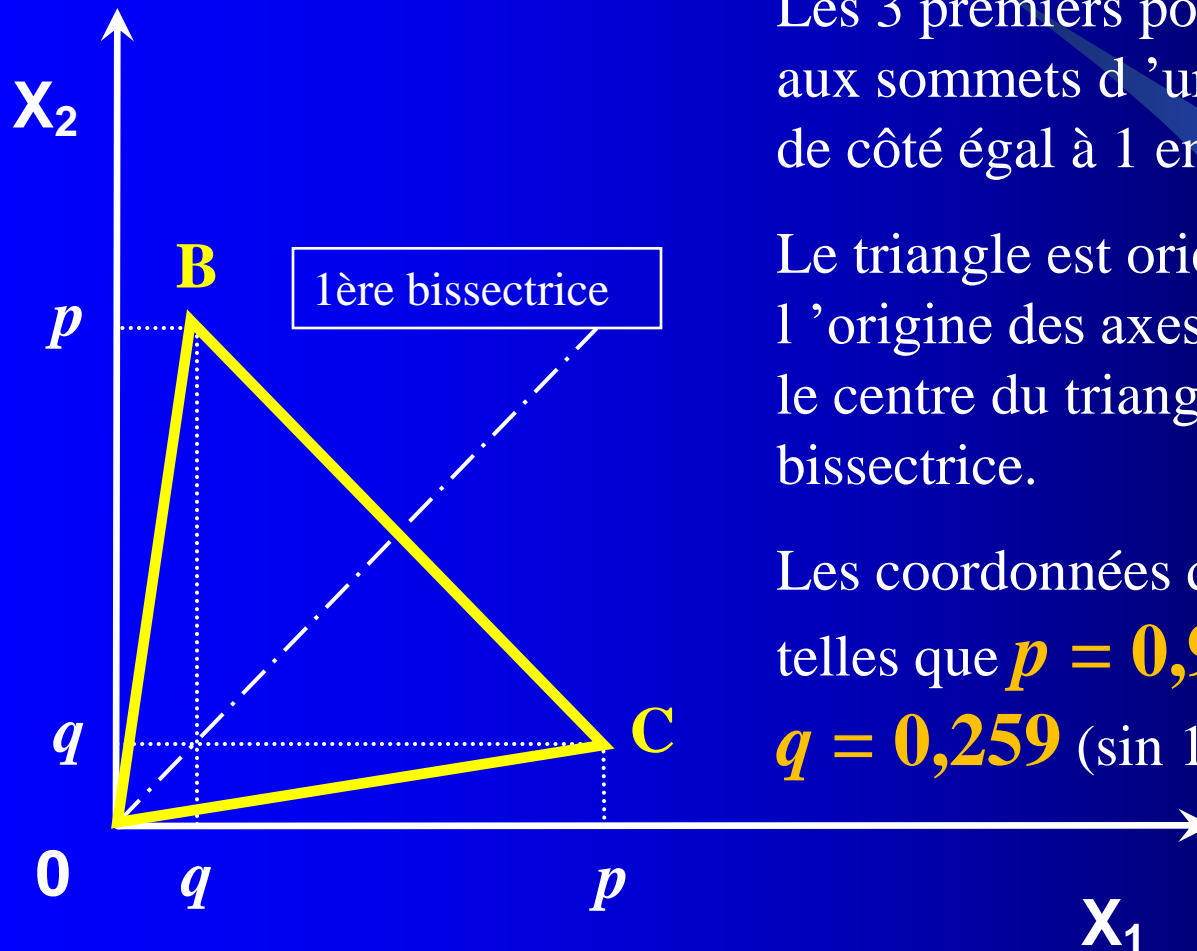
- Soient x_{10} et x_{20} les valeurs réelles des niveaux des facteurs au point de départ.
- Δx_1 et Δx_2 les pas de variation des facteurs choisis pour faire progresser les variables

Les valeurs réduites X_1 et X_2 des 2 facteurs utilisées dans le simplexe initial sont liées aux valeurs réelles par :

$$X_1 = \frac{x_1 - x_{10}}{\Delta x_1} \quad X_2 = \frac{x_2 - x_{20}}{\Delta x_2}$$

Les coordonnées des 3 points du simplexe initial sont : $(0, 0)$, (q, p) et (p, q) comme le montre le schéma suivant :

Le simplex initial dans le cas de 2 variables X_1 et X_2

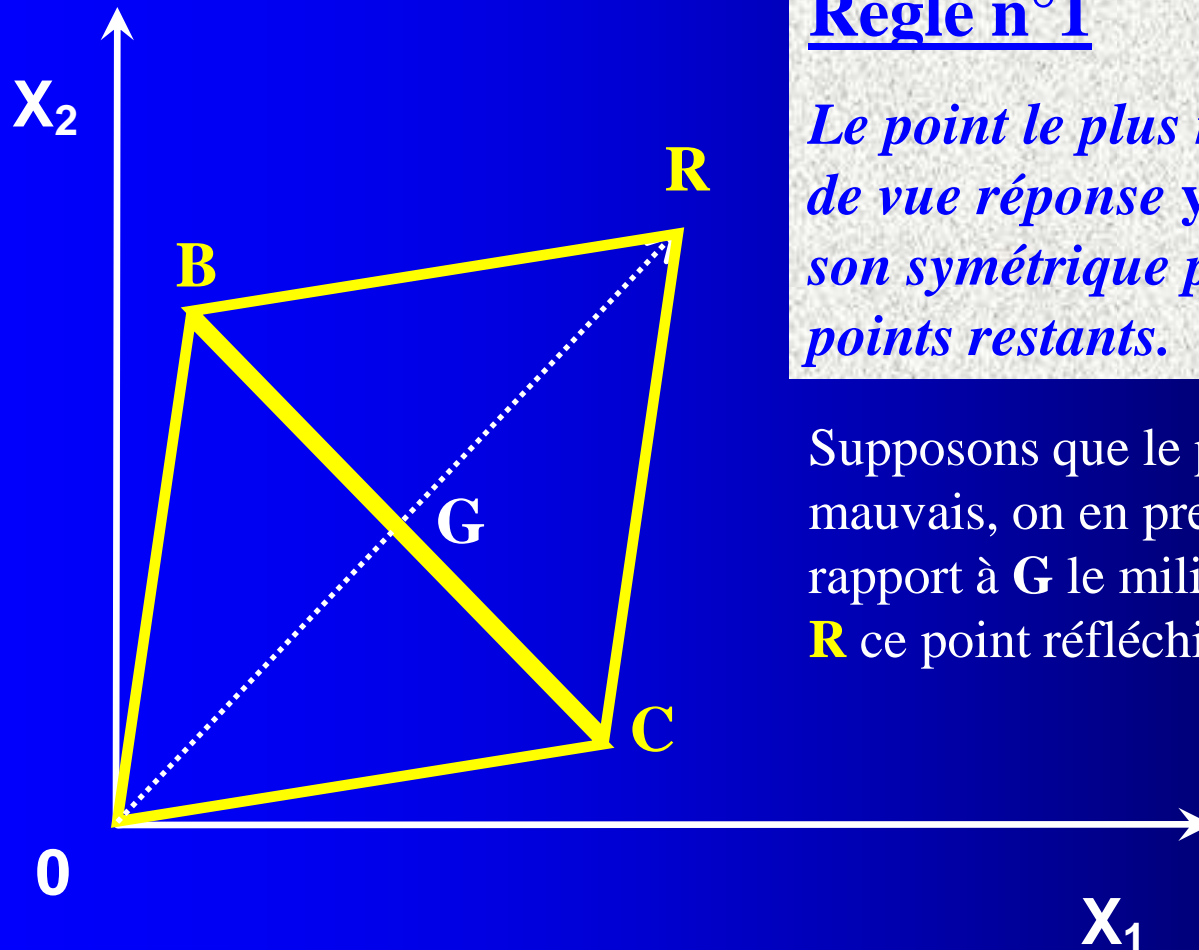


Les 3 premiers points expérimentés sont aux sommets d'un triangle équilatéral de côté égal à 1 en variables réduites.

Le triangle est orienté de telle sorte que l'origine des axes est un sommet et que le centre du triangle soit sur la 1ère bissectrice.

Les coordonnées des points B et C sont telles que $p = 0,966$ ($\cos 15^\circ$) et $q = 0,259$ ($\sin 15^\circ$)

Règles d'évolution du simplexe



Règle n°1

Le point le plus mauvais (du point de vue réponse y) est remplacé par son symétrique par rapport aux 2 points restants.

Supposons que le point 0 soit le plus mauvais, on en prend le symétrique par rapport à G le milieu de BC . On notera R ce point réfléchi.

Le nouveau simplexe à considérer est $[B,C,R]$

Dans ce nouveau simplex de progression vers l'optimum, il n'y a qu'une mesure de y à effectuer, celle de **R**. Il faut donc :

- calculer ses coordonnées réduites X_1 et X_2 .
- en déduire les valeurs x_1 et x_2 des 2 variables opératoires en unités de chaque grandeur.
- effectuer l'expérience et mesurer y_R .

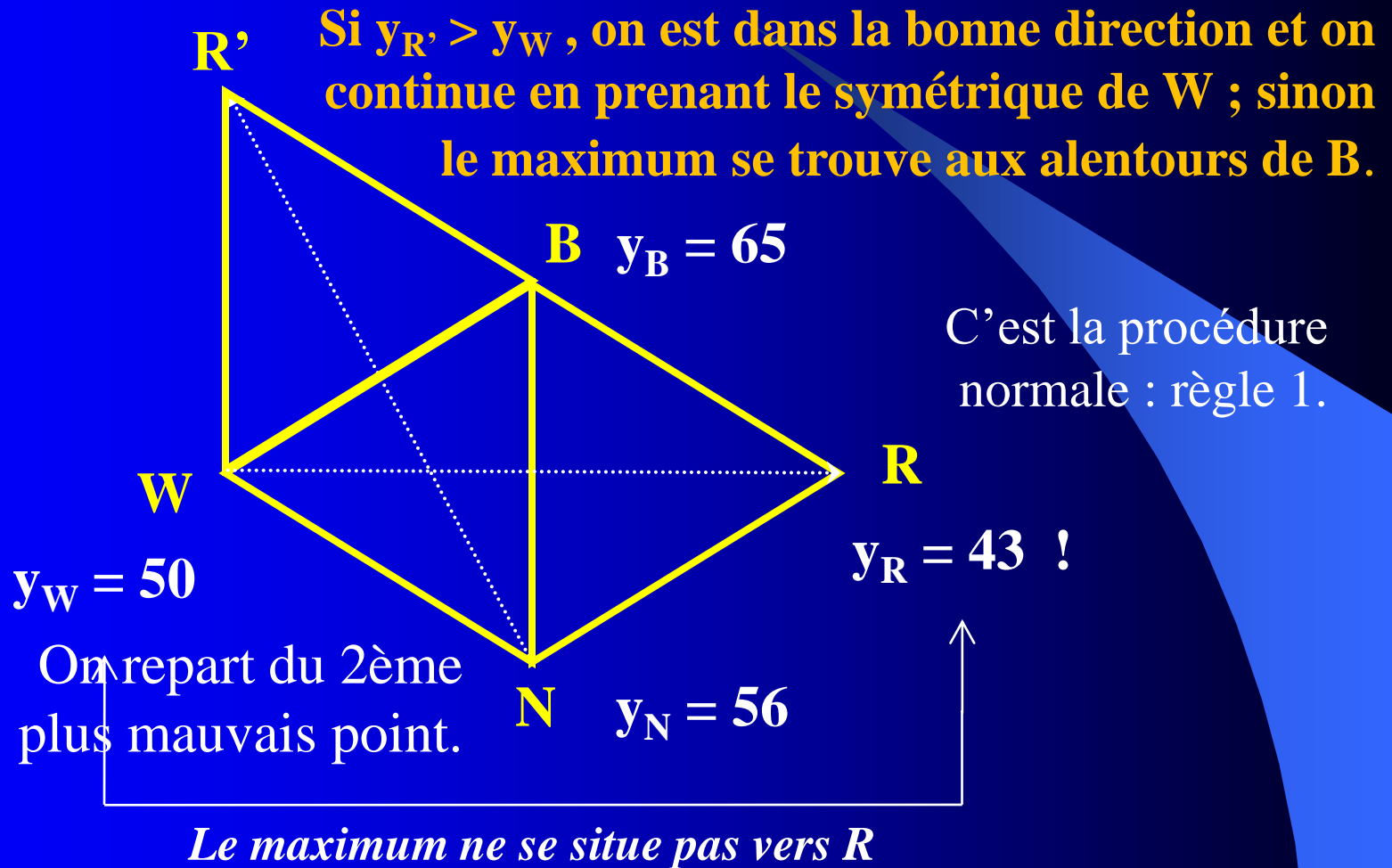
Le simplex [B,C,R] peut alors être analysé de la même façon que le premier, en comparant ses 3 réponses.

Cette opération est répétée et conduit à des simplex successifs en recherchant à chaque fois la plus mauvaise réponse.

Une exception à la règle 1 :

- Dans l'application de cette règle 1 de progression vers l'optimum, *il faut aussi comparer la réponse y_R à celle y_W de son symétrique W du plus mauvais point dans le simplex précédent* (W signifie mauvais, en anglais worst).
- **Lorsque y_R est plus mauvais que y_W** , il est évident que la direction $W \rightarrow R$ n'est pas une bonne direction pour trouver l'optimum... Il faut repartir du 2ème plus mauvais point (noté N) du simplex précédent et prendre son symétrique par rapport à G milieu de WB comme l'indique le schéma suivant.

Illustration : dans la recherche du maximum de la réponse y , le simplex actuel est $[W,B,N]$ et les 3 réponses sont connues.



Règles d'évolution du simplex (suite)

Règle n°2

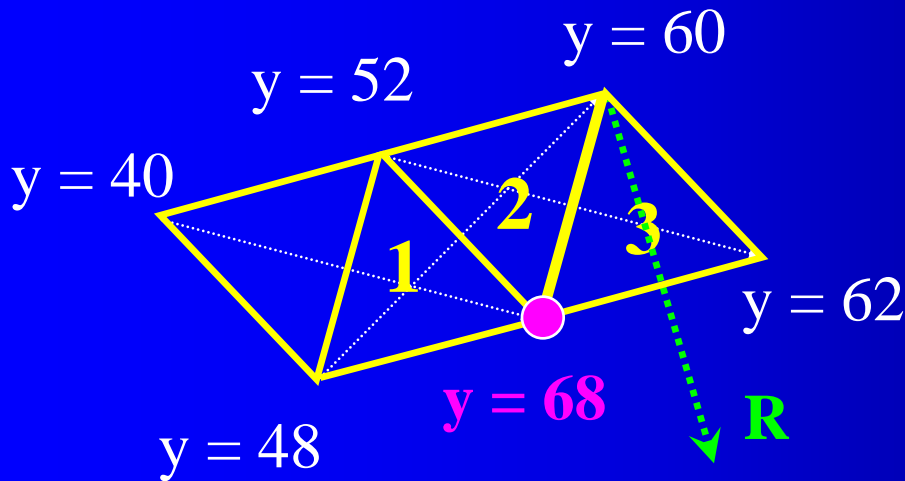
Lorsqu'un même point apparaît dans $k + 1$ simplex successifs sans être éliminé (les simplex tournent autour de ce point pivot), cela signifie que, soit la réponse en ce point est entachée d'une erreur de mesure, soit que ce point pivot est proche d'un optimum.

Il faut donc effectuer à nouveau l'expérience en ce point ; si la mesure est confirmée, la procédure normale continue. Sinon, il faut reprendre les simplex précédents pour redéfinir l'orientation.

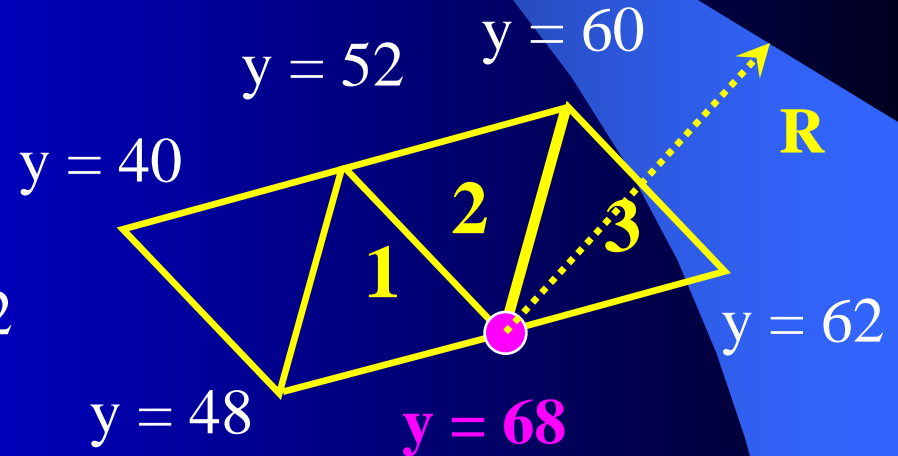
Illustration de la règle 2

Après 3 simplex ($k = 2$), le point rose n'est pas encore éliminé : une 2ème mesure de ce point s'impose.

Le point rose avait été correctement mesuré : l'évolution des simplex continue selon la direction R .



Le point rose a été surestimé (55 et non 68). Les simplex évoluent selon la direction R .



Règles d'évolution du simplex (fin)

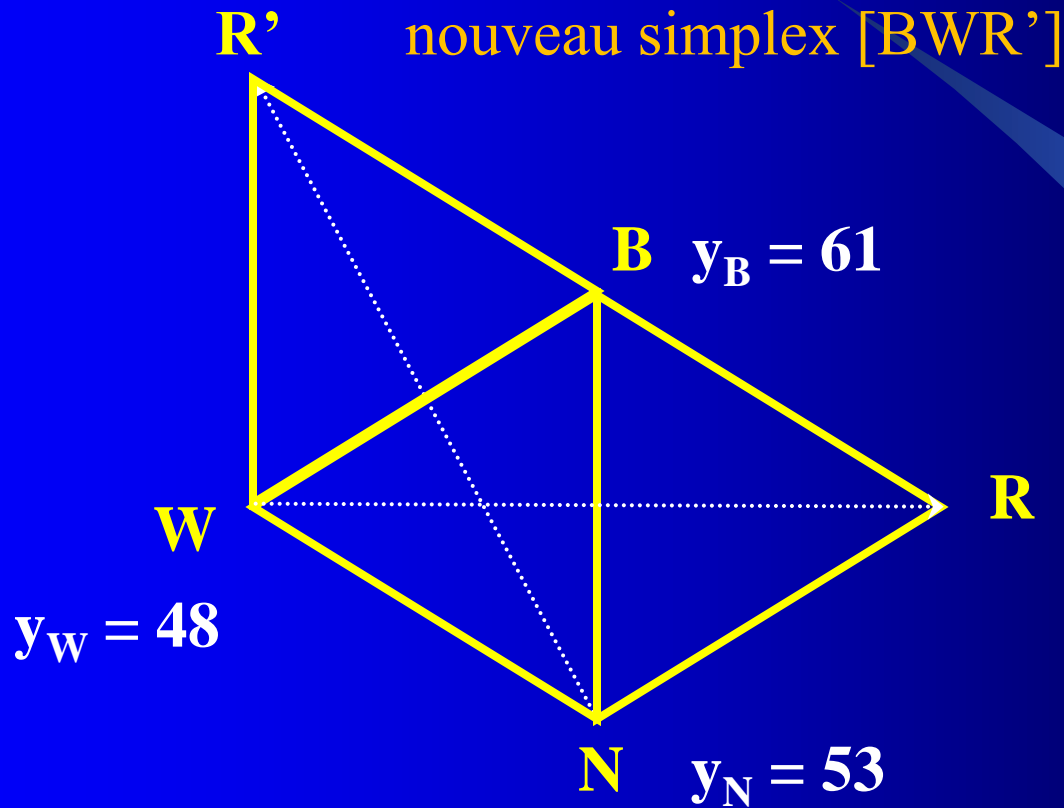
Règle n°3

Lorsque le symétrique du plus mauvais point dans un simplex est également le plus mauvais point dans le simplex suivant,

⇒ on prend le symétrique du 2ème plus mauvais point.

Remarque. *Cette règle est également appliquée lorsqu'il existe des contraintes expérimentales : le calcul du point réfléchi peut en effet conduire à des conditions expérimentales irréalisables en pratique. Une concentration a une limite (solubilité), une température peut en avoir une (décomposition du produit) ...etc.*

Illustration de la règle 3 : le simplex actuel est [W,B,N] et les 3 réponses sont connues.



impossible
d'expérimenter
au point **R** !

*Les conditions
sont
irréalissables*

On repart du 2ème
plus mauvais point.

Arrêt de la procédure des simplex

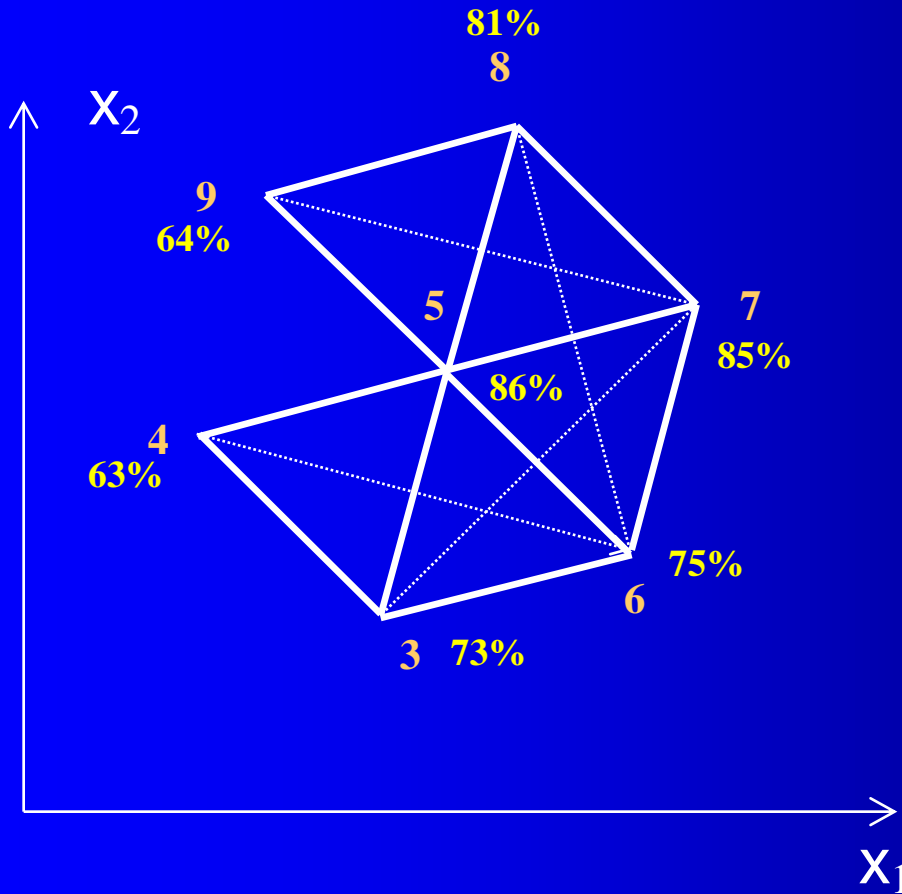
- La progression est arrêtée lorsque 6 simplex successifs autour d'un même point pivot forment un hexagone : l'optimum se trouve à l'intérieur de l'hexagone.



- Soit on considère que les réponses sont suffisamment voisines pour que la zone caractérise l'optimum.
- Soit on cherche à le cerner mieux en redémarrant une procédure de simplex avec un pas plus petit (diminution de la moitié ou du quart), soit on fait un plan composite centré sur le domaine hexagonal.



Illustration :



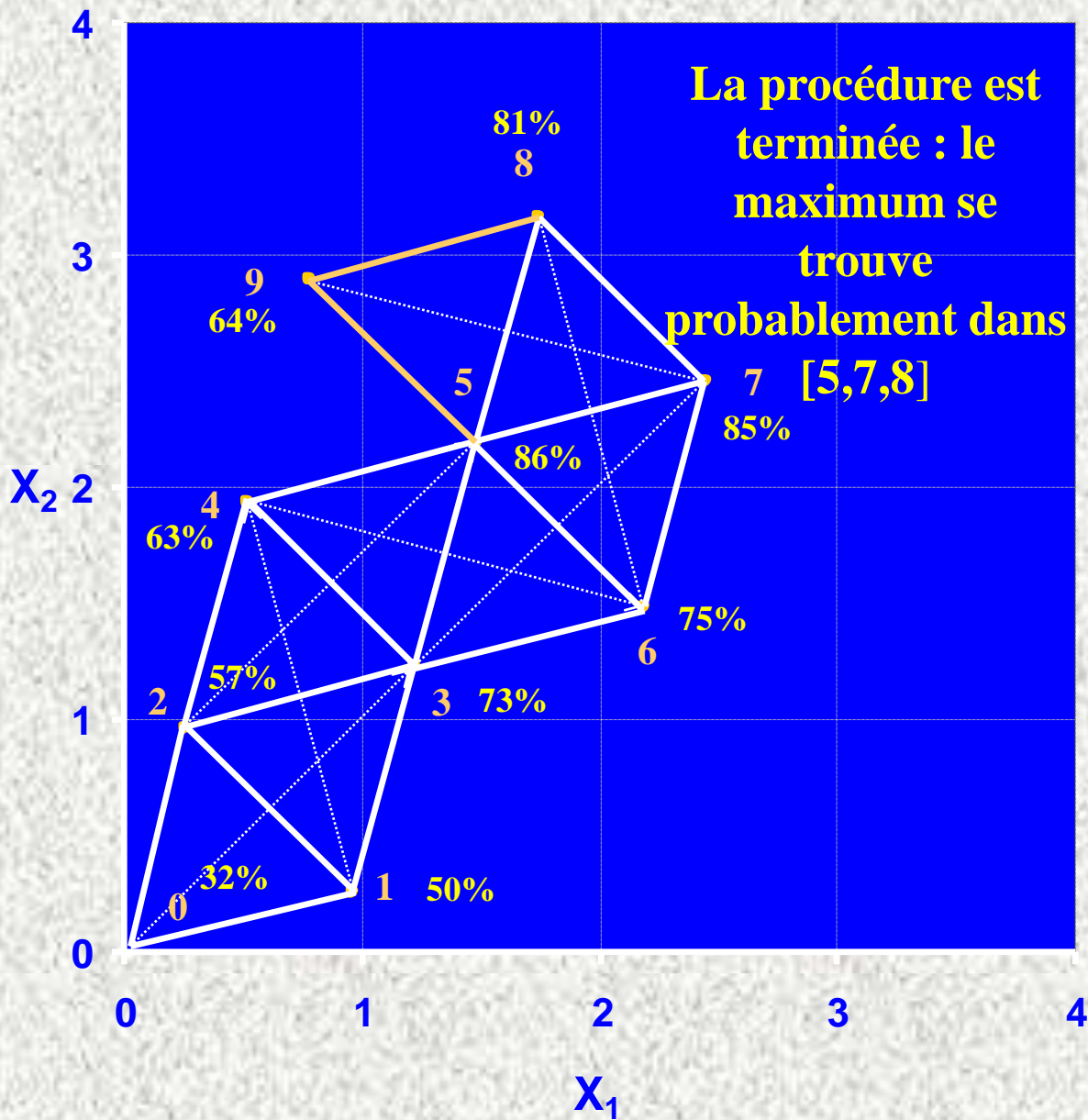
La procédure est terminée :
le simplex suivant consisterait à prendre le symétrique de 8 (règle 3) et on retomberait sur le point 4 déjà expérimenté. L'hexagone se refermerait.

Et ensuite ?



Exemple d 'application

- Exercice déjà étudié à propos des plans composites centrés (II - 5) concernant le rendement y de réaction en fonction de 2 variables opératoires :
 - le rapport molaire de $\text{Et}_3\text{N} / \text{M}_1$ (noté x_1)
 - le rapport molaire de M_2 / M_1 (noté x_2)
- Les calculs des conditions opératoires en valeurs réelles pour les essais successifs sont détaillés dans le polycopié.



1er simplex : le point 0

2ème simplex : le point

3ème simplex : le point

4ème simplex : le point

5ème simplex : le point

6ème simplex : le point

7ème simplex : le nouveau point 8 est le plus mauvais: le 2ème plus mauvais (7) est remplacé par son symétrique (le point 9)

Règle 3.

faut vérifier ce point 5,

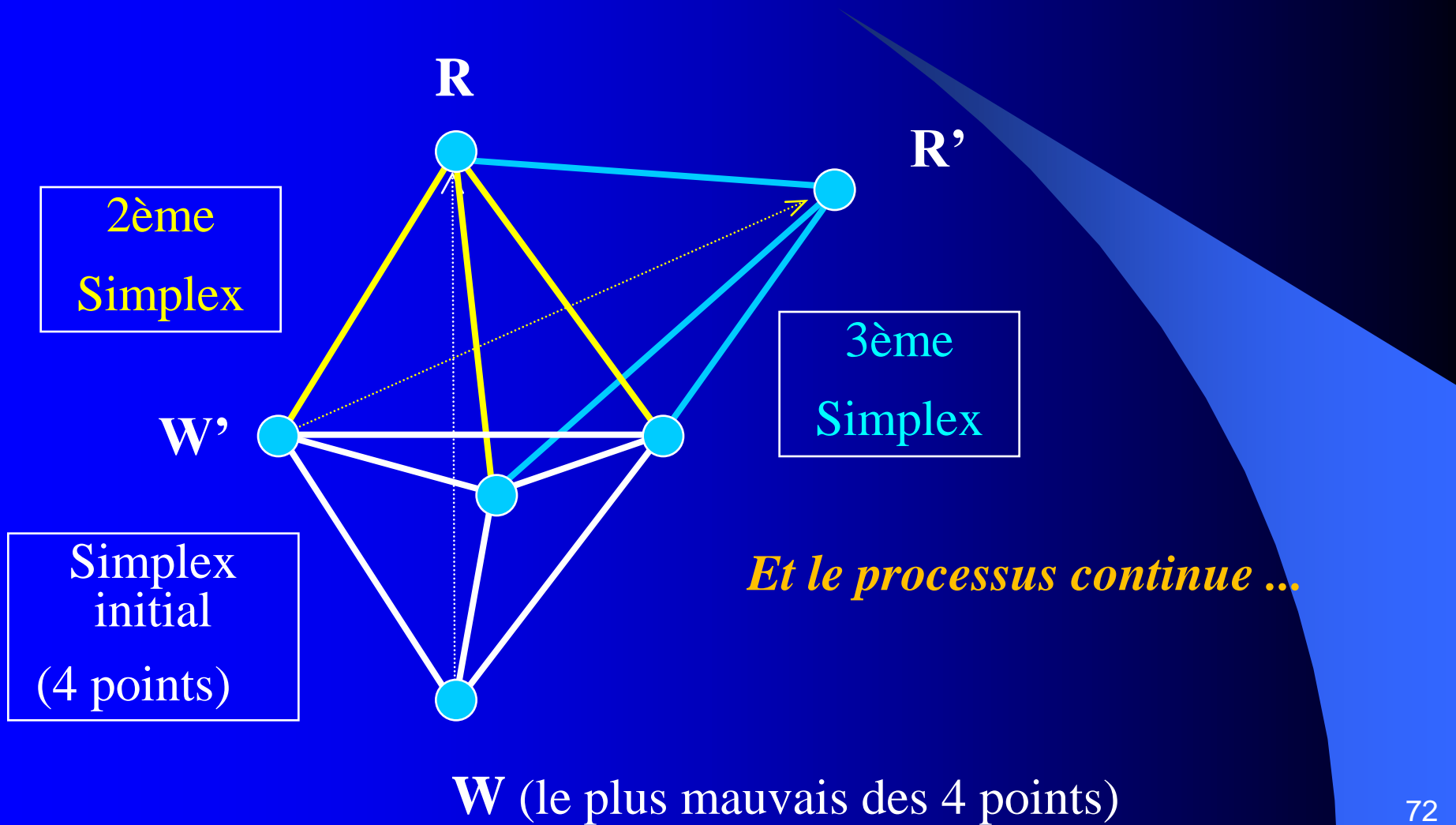
Règle 2.

La mesure donne à nouveau $y = 86\%$: le point était correct. La procédure continue

III - 3 : la méthode du simplex dans le cas de plus de 2 facteurs

- La méthodologie est généralisable lorsqu'on a plus de 2 facteurs intervenant dans l'optimisation. La stratégie est la même : mêmes règles de progression, calculs analogues à ceux vus à propos de 2 facteurs.
- Simplement le réseau de points qui forme un simplexe est plus complexe et les coordonnées réduites p et q sont différentes de 0,966 et 0,259 et fonction du nombre k de facteurs. Par exemple, dans un simplexe à 3 variables opératoires, le simplexe est un tétraèdre régulier pour lequel $p = 0,943$ et $q = 0,236$, illustré ci après :

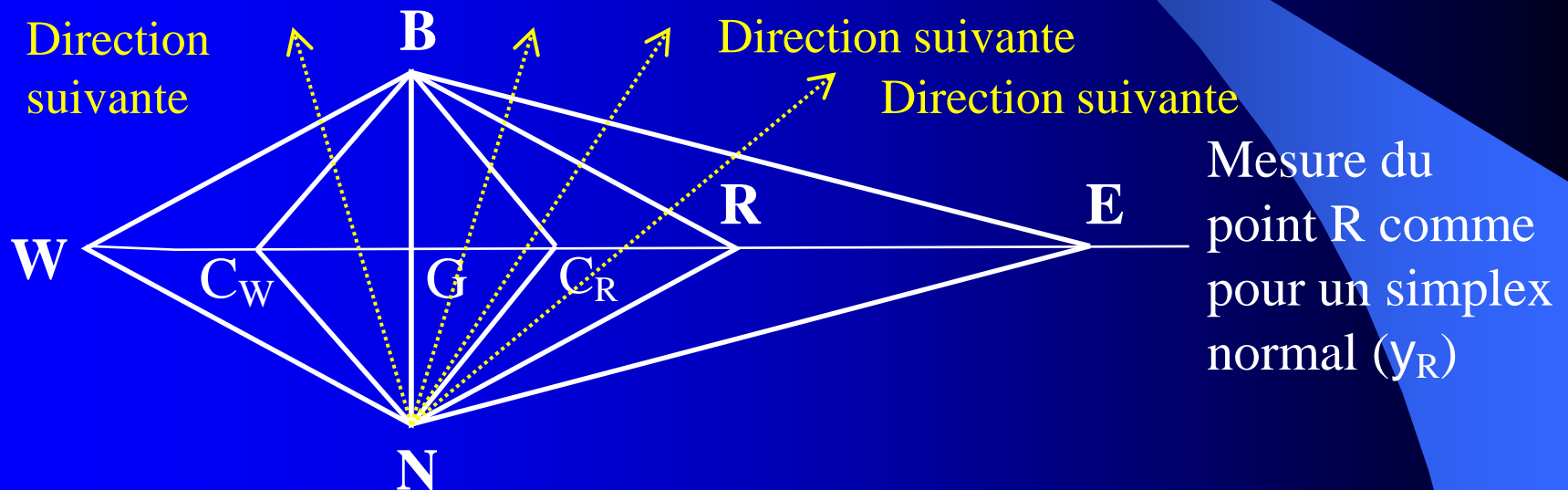
Évolution du simplex dans l'espace à 3 dimensions



III - 4 : le simplex modifié (Nelder et Mead)

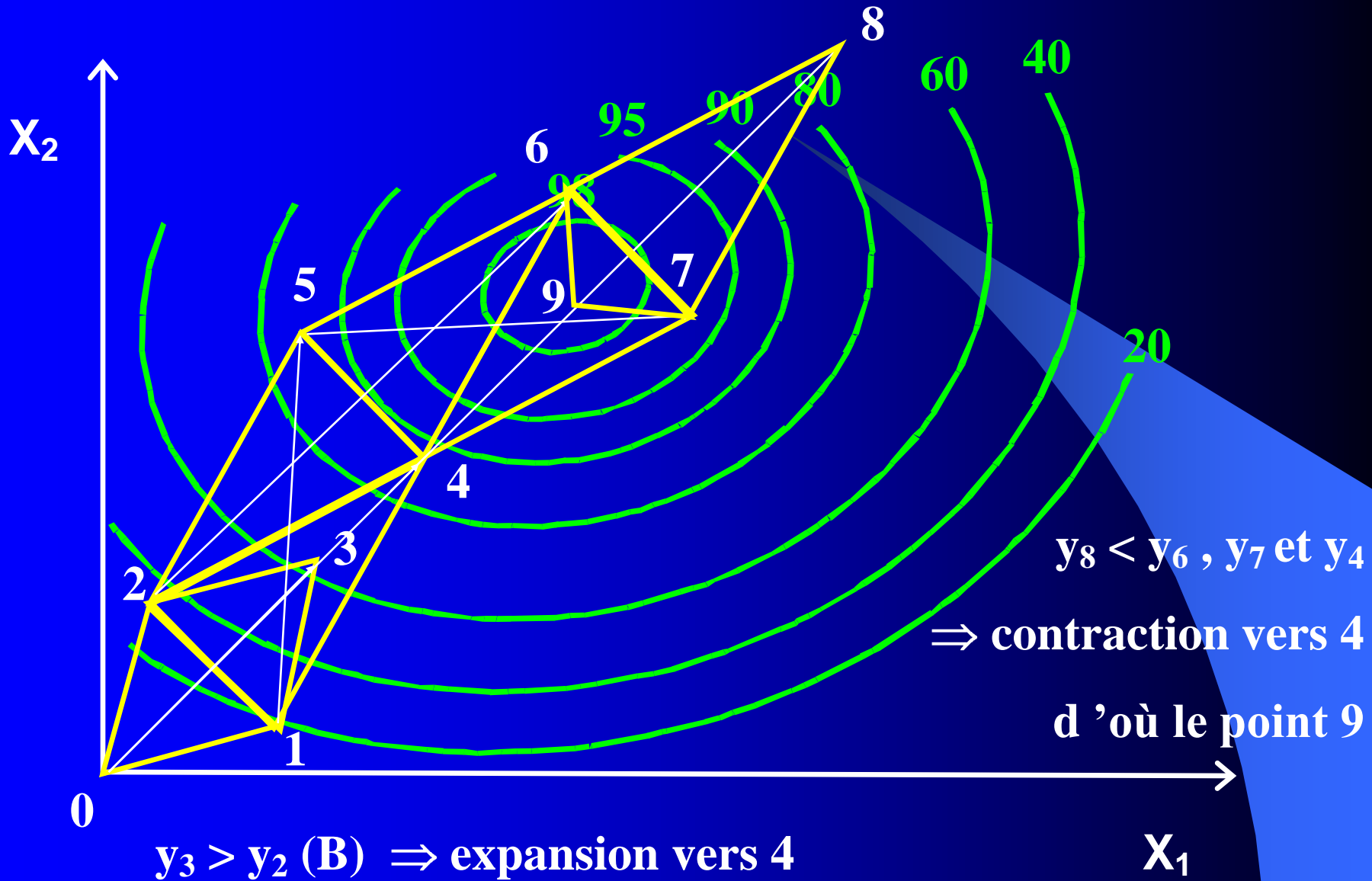
- Le simplex à pas fixes nécessite de bien choisir son pas ; s'il est trop petit, il faudra de nombreuses expériences pour obtenir l'optimum ; si le pas est trop grand, la localisation de l'optimum est peu précise.
- Pour remédier à cet inconvénient, Nelder et Mead (1965) ont proposé **un simplex à pas variable** où il intervient en plus de la **réflexion R**, la possibilité d'**expansion E** et de **contraction C** selon les valeurs des réponses.

- Si $y_R > y_B \rightarrow$ **expansion** (E) et simplex suivant : [N,B,E]
- Si $y_R < y_B$ et $y_R > y_N \rightarrow$ simplex suivant normal : [N,B,R]
- Si $y_R < y_N$ et $y_R > y_W \rightarrow$ **contraction** (C_R) et on a : [N,B, C_R]
- Si $y_R < y_N$ et $y_R < y_W \rightarrow$ **contraction** (C_W) et on a : [N,B, C_W]



W : point donnant la plus mauvaise réponse (Worst)
N : le second plus mauvais point (Next Worst)
B : point donnant la meilleure réponse (Best)

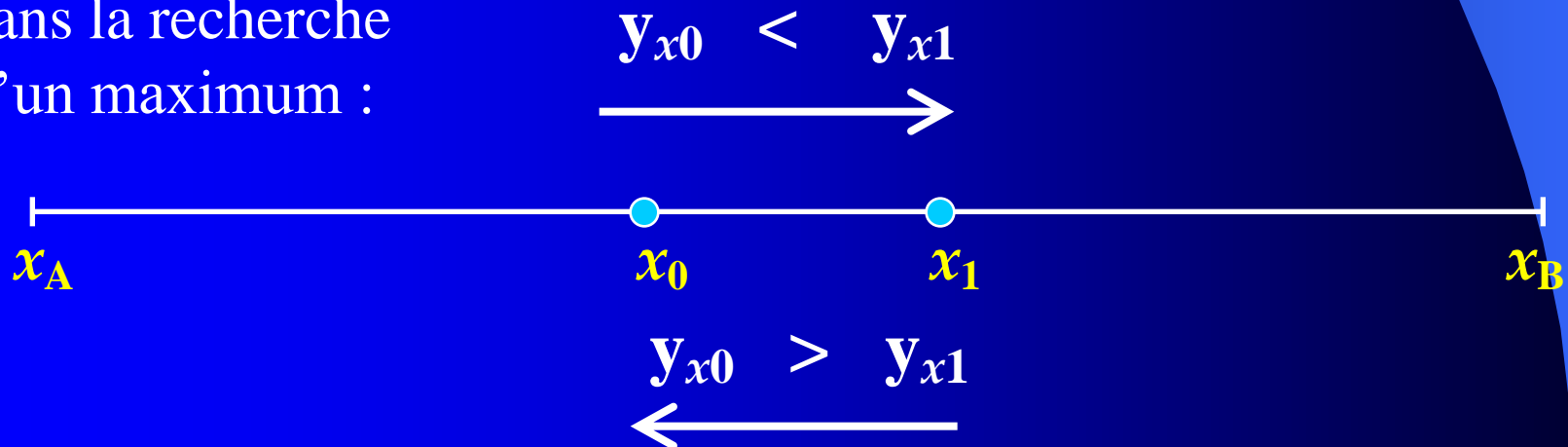
Illustration du fonctionnement du simplex à pas variable



III - 5 : optimisation dans le cas d'un seul facteur - méthode Uniplex

- Dans le domaine expérimental $[x_A, x_B]$ de la variable opératoire x dont dépend la réponse y , il s'agit de localiser la valeur x_M donnant l'optimum de y .
- Au départ, il faut réaliser 2 essais situés vers le milieu du domaine, pas trop rapprochés l'un de l'autre pour donner 2 réponses différentes, ce qui permet de savoir dans quel sens se diriger.

Dans la recherche d'un maximum :



- Divers algorithmes ont été proposés : **méthodes du nombre d'or**, **méthodes de Fibonacci** qui explorent l'intervalle $[x_A, x_B]$ par domaines successifs selon un protocole défini à l'avance.

Un écueil pour ces 2 méthodes est constitué par la qualité de la mesure de la réponse y : une erreur expérimentale risque d'inverser la direction supposée de l'optimum et ceci définitivement puisque l'évolution est fixée à l'avance.

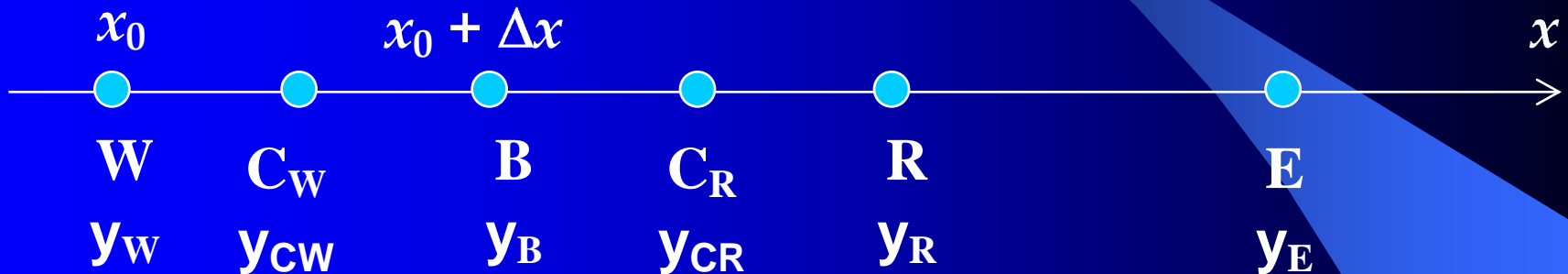
- Il est souvent préférable d'utiliser **la méthode Uniplex** qui ne présente pas cet inconvénient. Cette méthode correspond à celle du simplex pour lequel $k = 1$: le simplex initial est constitué par 2 points et lors de la progression vers l'optimum, il y a possibilité d'expansion et de contraction en plus de la réflexion des points (simplex modifié).

Règles de progression de la méthode uniplex

Choix de x_0 et Δx
 \Rightarrow simplex initial : 2 points

Mesure des 2 points
 W le plus mauvais
 B le meilleur ($y_B > y_W$)

Réflexion : R
 symétrique de W
 par rapport à B



- ① Si $y_R > y_B \rightarrow$ **expansion** vers E. Simplex suivant : **BE**
- ② Si $y_R < y_B$ et $y_R > y_W \rightarrow$ **contraction** vers B (C_R). Simplex suivant : **BC_R** si $y_{C_R} > y_R$ ou **BR** si $y_{C_R} < y_R$
- ③ Si $y_R < y_W \rightarrow$ **contraction** vers W (C_W). Simplex suivant : **BC_W** si $y_{C_R} > y_W$. Sinon vérifier la mesure y_B .

Exemple d'application

Optimisation de la fraction molaire x d'un constituant pour obtenir le rendement de réaction y % le plus élevé possible. *L'incertitude sur une mesure de rendement est de 2%.*

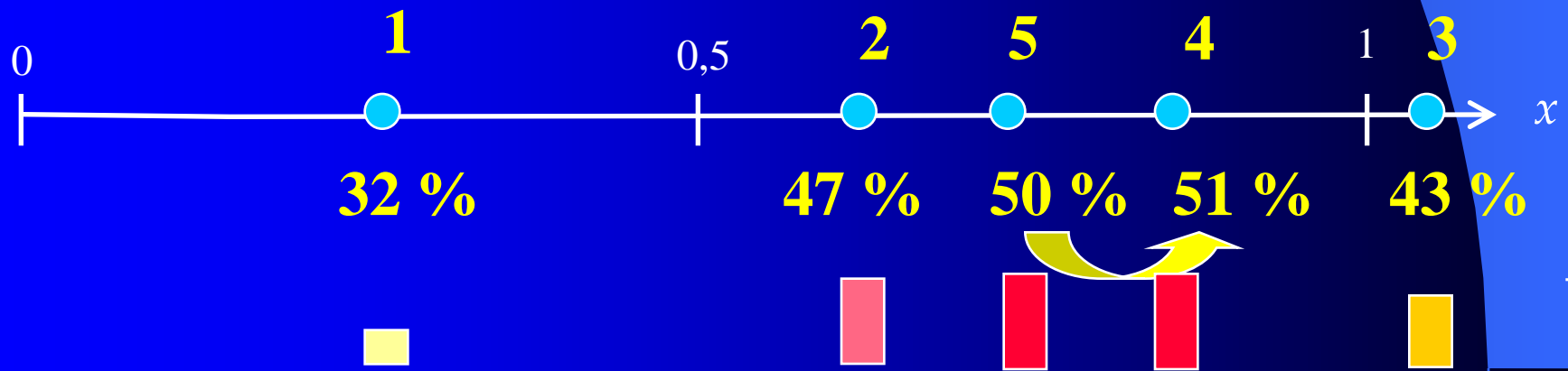
On part de $x_1 = 0,25$ avec un pas $\Delta x = 0,40$; on a donc $x_2 = 0,65$

Le 3ème point a pour fraction molaire $x_3 = 1,05$ ($0,65 + 0,40$)

Comme $y_3 < y_2$ et $y_3 > y_1$, on a $x_4 = 0,85$ (milieu de 2-3)

Le point suivant est $x_5 = 0,75$ (milieu de 2-4)

L'optimisation est terminée : les rendements ne diffèrent plus.



- *Une remarque importante*

Le choix d'une fraction molaire x_1 comprise entre 0,75 et 0,85 assure le rendement de réaction le plus élevé (environ 50 %), mais il faut ajouter que l'optimum déterminé ne concerne que ce facteur ... pour les conditions expérimentales choisies.

Car cet optimum déterminé n'a d'intérêt pratique que si ce facteur est seul à agir sur le rendement y ou s'il n'existe pas d'interaction(s) entre les facteurs.

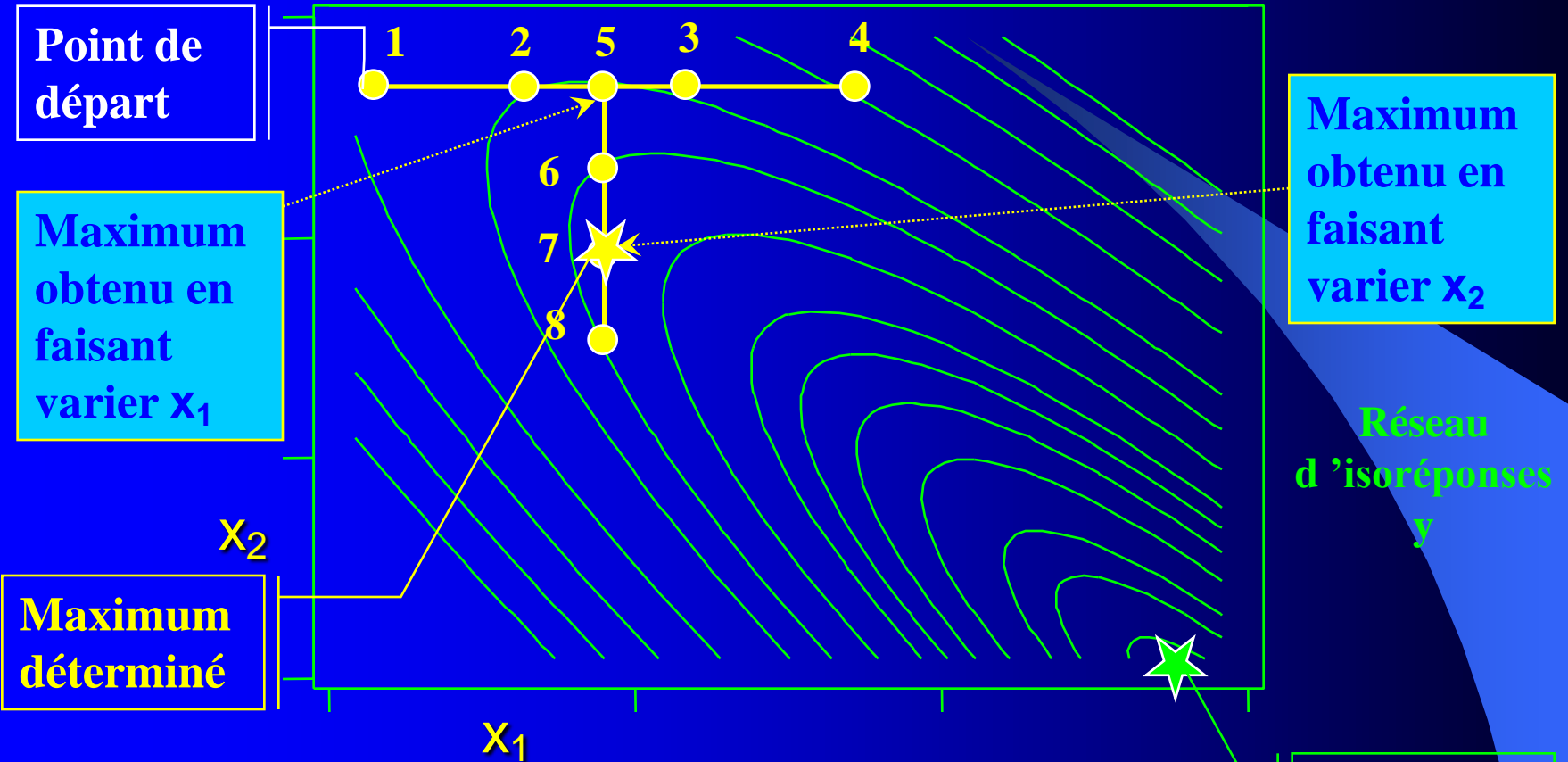
Dans le cas contraire, les autres facteurs doivent être maintenus constants lors des essais successifs ; ajoutons qu'on ne peut pas préjuger de la valeur optimale de ce facteur quand on fera varier les autres facteurs.

III -6 : La méthode « 1 seul facteur variable à la fois »

- inventée par FRIEDMAN et SAVAGE, cette méthode intuitive consiste à faire varier un facteur à la fois, à le fixer à sa valeur optimale puis à rechercher successivement l'optimum pour les autres facteurs. **Pour 2 facteurs et en 2 étapes**, cette méthode ne permet pas d'atteindre l'optimum s'il existe une interaction entre les facteurs, comme le montre le schéma suivant :

1ère phase : variation de x_1 à x_2 constant

2ème phase : variation de x_2 à x_1 constant



*Le maximum déterminé est très loin du max vrai ...
Ce n'est pas une procédure à recommander.*

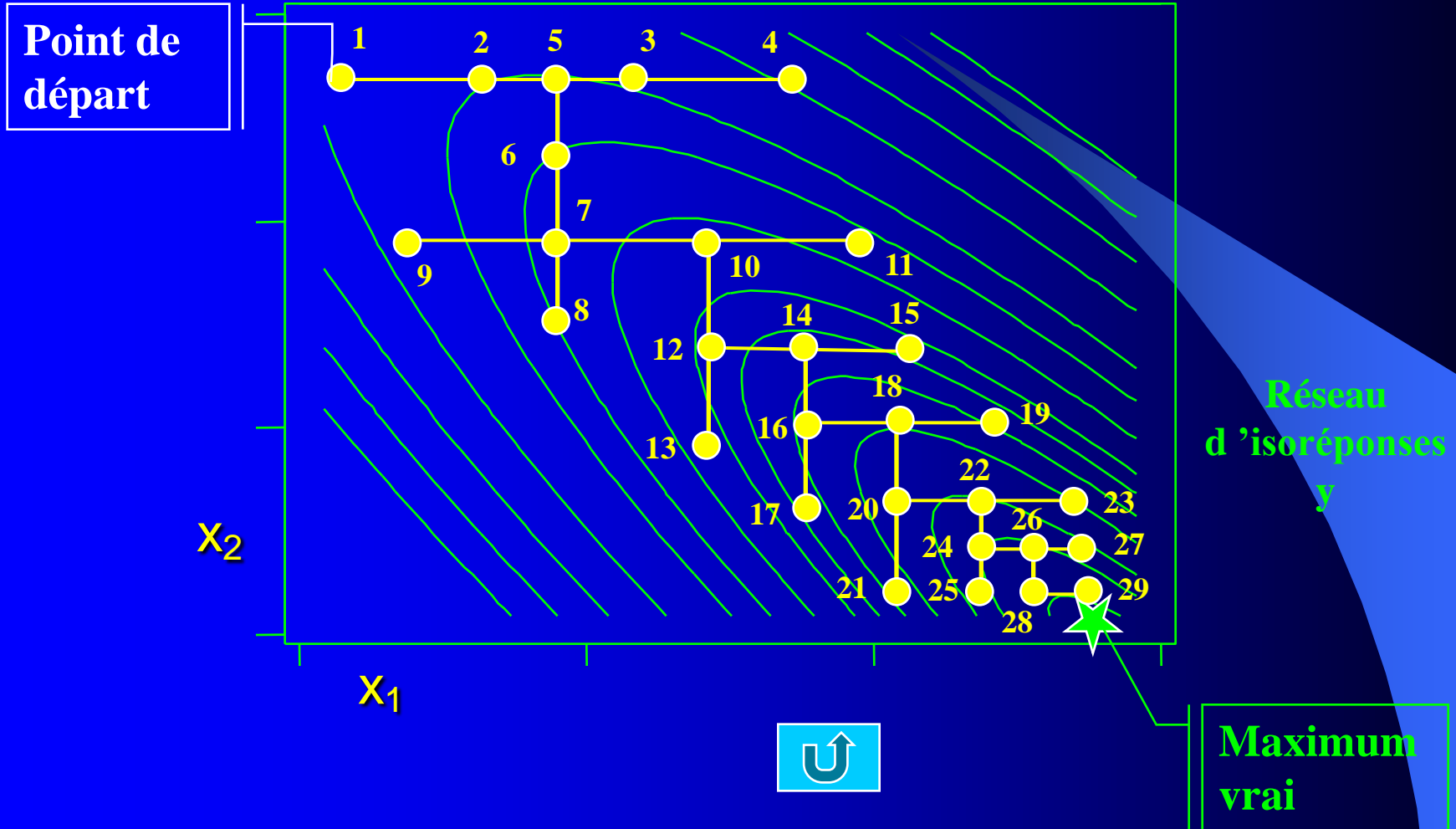
- Pour que la méthode de **Friedman** soit applicable en pratique et donne confiance dans l'optimum déterminé, il est nécessaire après 2 étapes de refaire varier x_1 pour rechercher un nouvel optimum, puis à nouveau x_2 et réitérer ce processus autant de fois qu'il le faut jusqu'à ce que la réponse n'augmente plus...

Cette méthode présente l'inconvénient de réclamer un grand nombre d'essais, comme l'illustre le schéma suivant

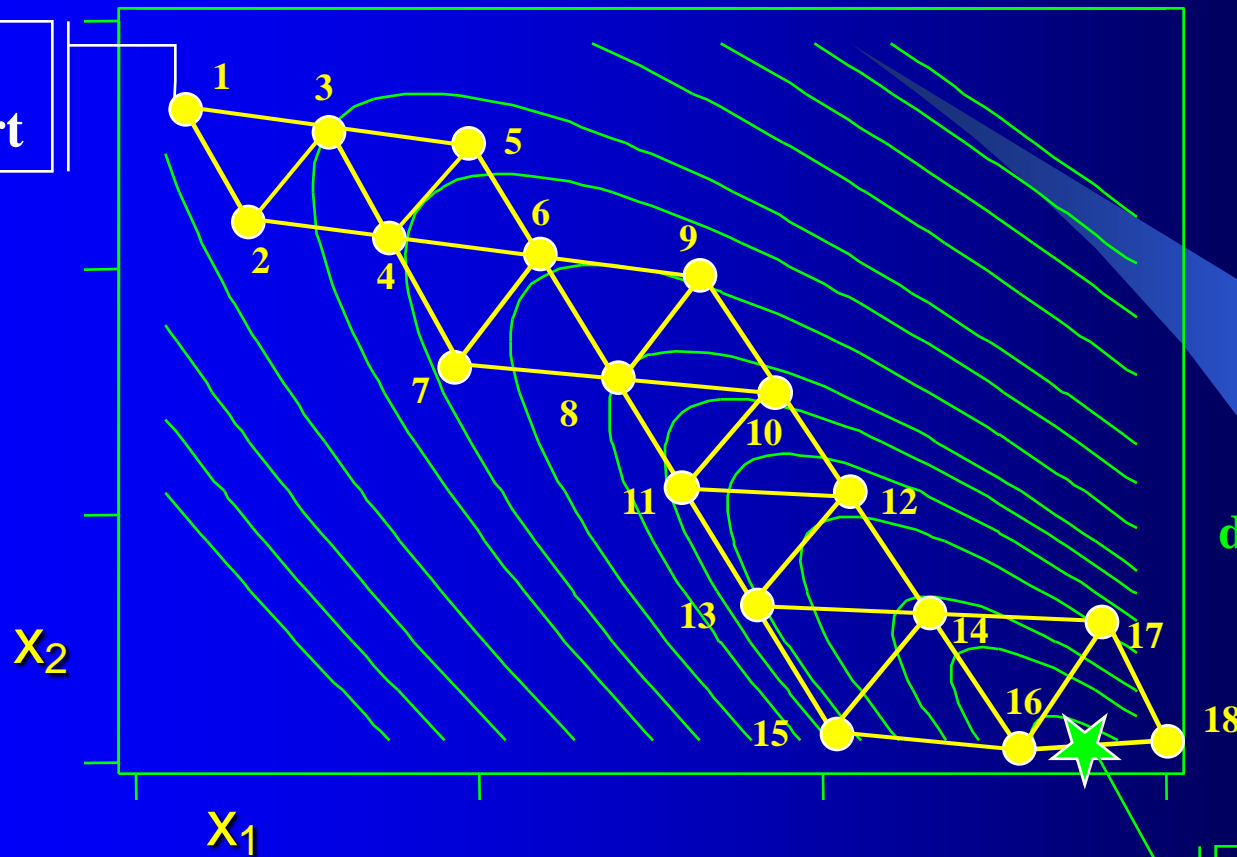
De plus, il est des cas particuliers où cette technique ne permet pas d'atteindre l'optimum vrai (blocage sur une arête de la surface de réponse).

- Il est très préférable d'utiliser la méthode du simplexe qui nécessite moins d'essais et ne peut pas présenter ce phénomène de blocage.

Continuons le *recherche* expériences (29) pour parvenir à *maximum* vrai avec la méthode de Friedman ...



Simplex
de départ



Réseau
d'isoréponses
 y

Maximum
vrai

Conclusion : simplex ou plan d 'expérience ?

avantages

Plan d 'expérience

- mise en évidence de chaque facteur et des interactions
- nombre d 'essais déterminé (planification)
- possibilité de prédire des réponses
- applicable au cas où on a plusieurs réponses y

Simplex

- ne nécessite pas de modélisation
- applicable à n 'importe quelle surface de réponse
- détection rapide des points aberrants
- nécessite parfois peu d 'essais

Conclusion : simplex ou plan d 'expérience ?

inconvénients

Plan d 'expérience

- nécessite de choisir un modèle qui peut se révéler inadéquat
- difficulté de détecter des points aberrants
- localisation parfois peu précise de l 'optimum à cause de l 'approximation du modèle

Simplex

- le rôle de chaque facteur ne peut pas être mis en évidence
- le nombre d 'essais est au départ imprévisible
- difficile à appliquer à des essais de longue durée
- un simplex par réponse y