Plans optimaux pour le krigeage

Victor Picheny

29 janvier 2015

1. Rappels sur le krigeage

2. Plans optimaux

3. Plans adaptatifs

4. Plans d'expériences : conclusion et perspectives

Notation et propriétés

Hypothèse fondamentale

$$Y(\mathbf{x}) \sim \mathcal{PG}\left(\mu(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')\right)$$

On note :

- $m(\mathbf{x}^*) = \mathbb{E}(Y(\mathbf{x}^*)|\mathbf{Y}_n)$: moyenne du krigeage
- $s^2(\mathbf{x}^*) = var(Y(\mathbf{x}^*)|\mathbf{Y}_n)$: variance de prédiction

Aux points du plan :

$$m(\mathbf{x}_i) = y_i$$

$$s^2(\mathbf{x}_i) = 0$$

On a bien un interpolateur !

Krigeage avec tendance (krigeage universel)

Tendance : $\mu(\mathbf{x}) = \sum_{k=1}^{p} \beta_k f_k(\mathbf{x})$

Estimation de β

Moindres carrés généralisés :

$$\beta^* = \left(\mathbf{F}^T\mathbf{K}^{-1}\mathbf{F}\right)^{-1}\mathbf{F}^T\mathbf{K}^{-1}\mathbf{Y}_n$$

Equations :

$$\begin{aligned}
m_{UK}(\mathbf{x}^*) &= \mathbf{f}(\mathbf{x}^*)\beta^* + \mathbf{k}(\mathbf{x}^*)^T\mathbf{K}^{-1}\left(\mathbf{Y}_n - \mathbf{F}\beta^*\right) \\
s^2_{UK}(\mathbf{x}^*) &= \sigma^2 - \mathbf{k}(\mathbf{x}^*)^T\mathbf{K}^{-1}\mathbf{k}(\mathbf{x}^*) \\
&+ \left(\mathbf{f}(\mathbf{x}^*) - \mathbf{F}^T\mathbf{K}^{-1}\mathbf{k}(\mathbf{x}^*)\right)^T\left(\mathbf{F}^T\mathbf{K}^{-1}\mathbf{F}\right)^{-1}\left(\mathbf{f}(\mathbf{x}^*) - \mathbf{F}^T\mathbf{K}^{-1}\mathbf{k}(\mathbf{x}^*)\right)
\end{aligned}$$

NB : la variance ne dépend pas de la valeur des observations

Influence de la fonction de covariance

Covariance la plus répandue en krigeage : covariance gaussienne anisotrope.

$$k(\mathbf{x}, \mathbf{x}') = \sigma^2 \exp\left(\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{\theta^2}\right)$$

La fonction de covariance détermine une grande partie des caractéristiques du processus :

- son amplitude : *variance du processus* σ^2
- son "activité" : *portée* θ
  grande portée : processus régulier (plat) ;
  portée faible : processus chahuté
- autres : dérivabilité, etc.

Estimation des paramètres de covariance

Plusieurs techniques possibles :

- Validation croisée, variogramme (géostatistique)
- **Vraisemblance**

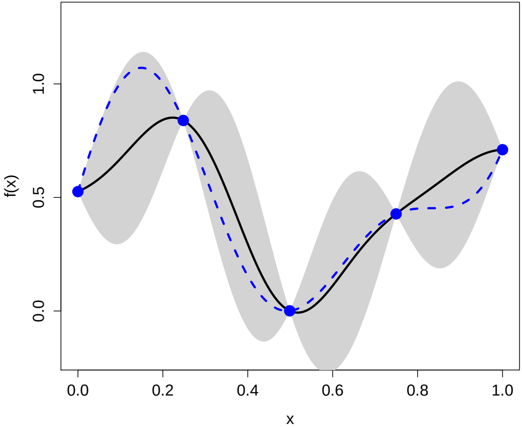
Maximiser la vraisemblance = maximiser la densité de proba. de $\mathbf{Y}_n$ = minimiser la log-vraisemblance :

$$l = \log \det \mathbf{K} + \mathbf{Y}^T \mathbf{K}^{-1} \mathbf{Y}_n$$

C'est le cas du package *DiceKriging*.

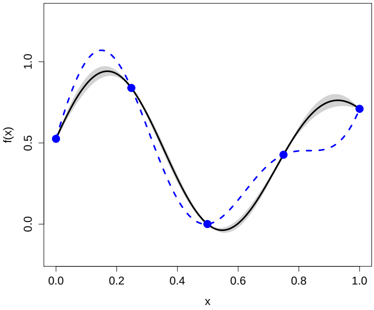
Effet de l'erreur de l'estimation des paramètres

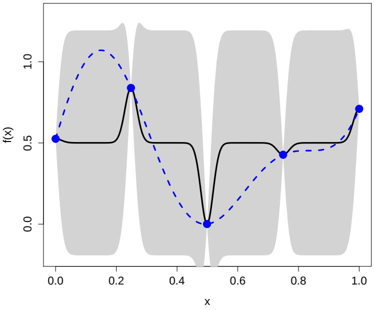
Modèle précis



Effet de l'erreur de l'estimation de la portée

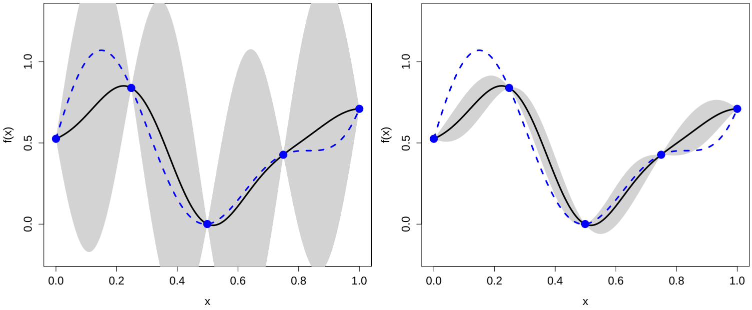
Sur-estimation de la portée / sous-estimation de la portée :





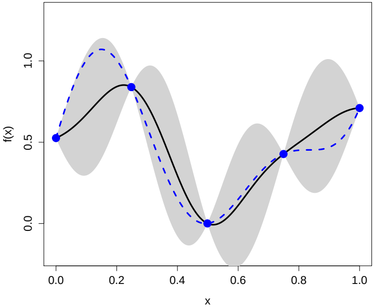
Effet de l'erreur de l'estimation de la variance

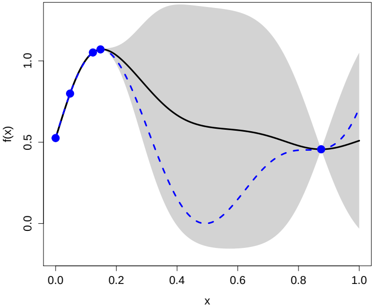
Sur-estimation de la variance / sous-estimation de la variance :



Effet de la répartition des points

Les deux modèles ont la même covariance et le même nombre de points :





Qualité d'un modèle

Deux types d'information

- Qualité de l'apprentissage des paramètres
- Qualité de la prédiction

Parallèle avec la régression

- Qualité de l'apprentissage des paramètres : D-optimalité
- Qualité de la prédiction : G-optimalité

Le théorème de D-G équivalence montre que les même plans optimisent l'apprentissage des paramètres et la prédiction !

Vrai pour le krigeage également ?

D-optimalité (krigeage universel)

- On considère que l'information importante est contenue dans les coefficients.
- Equivalent à la régression avec résidus corrélés

Moindres carrés généralisés :

$$\beta^* = \left(\mathbf{F}^T \mathbf{K}^{-1} \mathbf{F}\right)^{-1} \mathbf{F}^T \mathbf{K}^{-1} \mathbf{Y}$$

Critère à maximiser :

$$det \left(\mathbf{F}^T \mathbf{K}^{-1} \mathbf{F}\right)$$

Problème : dépend de la covariance $\mathbf{K}$.
Cf. régression classique

D-optimalité... pour la vraisemblance

- Objectif : meilleur apprentissage des coefficients de la covariance
- Quantité correspondante : log-vraisemblance :

$$l = \log\det \mathbf{K} + \mathbf{Y}^T \mathbf{K}^{-1} \mathbf{Y}$$

- Type de plans obtenus : plus grande variété des interdistances possibles
- Pas de résultats théoriques à ce jour

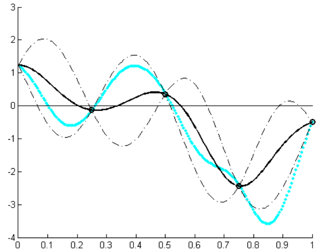
G-optimalité (1/2)

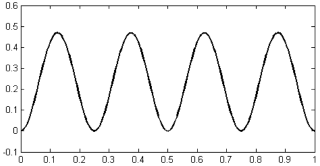
Hypothèse principale

Les paramètres de covariance sont connus avec précision.

Mesure de la qualité du modèle

Variance de prédiction





G-optimalité (2/2)

Plan G-optimal : minimise le maximum de la variance de prédiction :

$$\min_{\mathbf{x}_1,\dots,\mathbf{x}_n} G = \max_{\mathbf{x}^*} s_K^2(\mathbf{x}^*)$$

Quelques résultats

- en dimension 1 : grilles uniformes (pour certaines covariances)
- en général : plans space-filling
- tendance à remplir les bords du domaine

I-optimalité (1/2)

Critère : espérance de la variance de prédiction

$$\min_{\mathbf{x}_1,\dots,\mathbf{x}_n} I = E\left(s_K^2(\mathbf{x}^*)\right) = E\left[\left(m_K(\mathbf{x}) - Y(\mathbf{x})\right)^2\right]$$

Forme intégrale :

$$I = \int s_K^2(\mathbf{x}) d\mu(\mathbf{x})$$

Si μ est la mesure de Lebesgue :

$$I = \int_D s_K^2(\mathbf{x}) d\mathbf{x}$$

I-optimalité (1/2)

Critère I (ou IMSE)

Mesure globale de l'incertitude du modèle

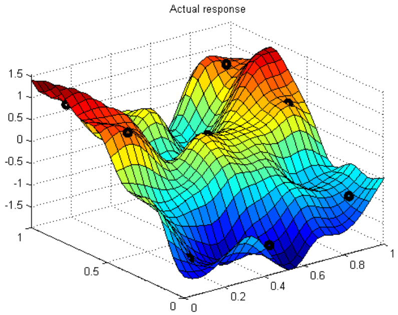
I dépend :

- du plan d'expérience
- de la fonction de covariance estimée
- de la tendance choisie

Type de plans I-optimaux

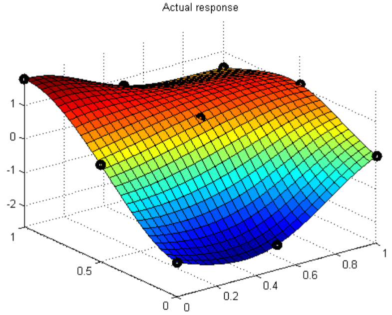
- Pas de résultats théoriques
- En général : remplissage d'espace

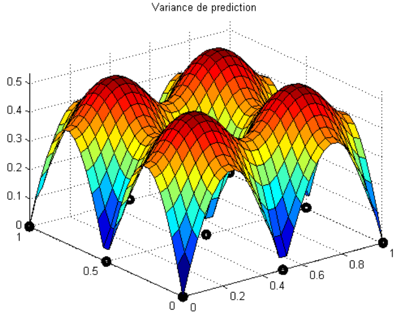
Vraie fonction



Prédiction avec un plan factoriel à 3 niveaux

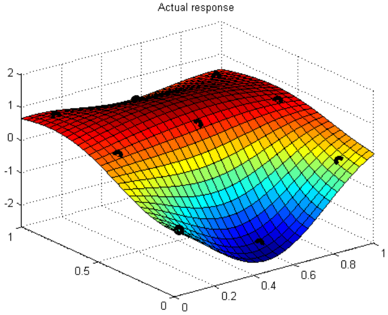
- IMSE = 0.3647
- maxMSE = 0.51

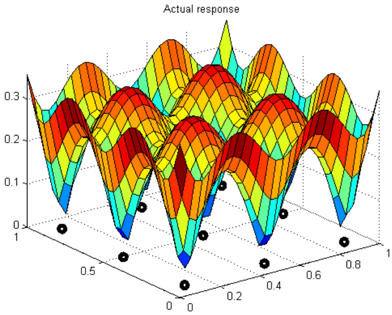




Prédiction avec un plan factoriel de côté 80%

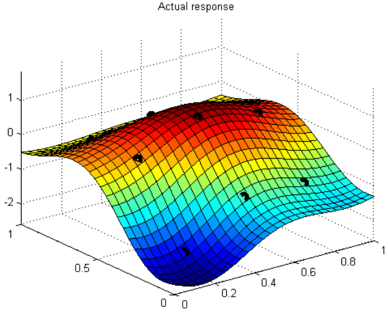
- IMSE = 0.2322
- maxMSE = 0.33

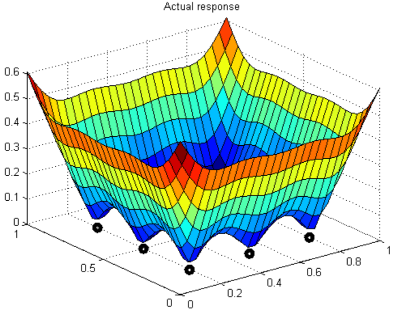




Prédiction avec un plan factoriel de côté 60%

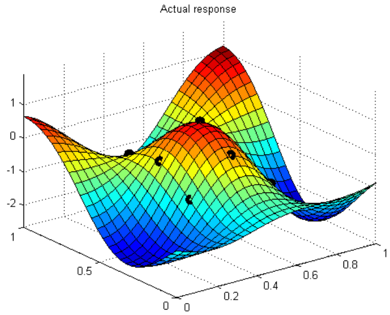
- IMSE = 0.2279
- maxMSE = 0.6

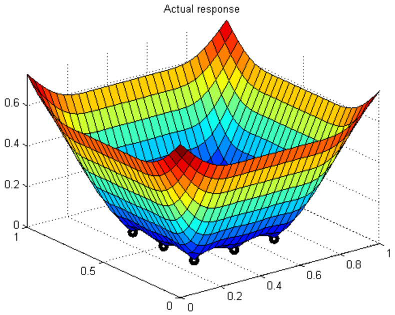




Prédiction avec un plan factoriel de côté 40%

- IMSE = 0.3005
- maxMSE = 0.75





Problèmes techniques (1/2)

G-optimalité

- Le calcul du critère nécessite une boucle d'optimisation interne
- Problème global d'optimisation insoluble : grande dimension, multimodal, ...

Solutions

- Optimisation pour une famille de plans (factoriels, LHS)
- Opération sur les plans : contraction, dilatation
- Construction séquentielle

Problèmes techniques (2/2)

I-optimalité

- Le calcul du critère nécessite un calcul d'intégrale numérique
- Solutions : grille dense, Monte-Carlo, suite à faible discrépance
- Quelques simplifications analytiques parfois

$\rightarrow$ Très coûteux à évaluer !

Solutions

- Optimisation pour une famille de plans (factoriels, LHS)
- Opération sur les plans : contraction, dilatation
- Construction séquentielle

Plans adaptatifs

Principe

- Remplacer un problème d'optimisation de dimension $n \times d$
- ... par n problèmes à d dimensions

Problème d'optimisation complet (C = maxMSE ou IMSE) :

$$\min C(\mathbf{x}_1, \ldots, \mathbf{x}_n)$$

Problème d'optimisation adaptatif (C = maxMSE ou IMSE) :

Pour i allant de 1 (ou $1 \leq k \leq n$) à n :

$$\min C_{\mathbf{x}_1, \ldots, \mathbf{x}_{i-1}}(\mathbf{x}_i)$$

Application aux critères G et I

Problème

- Pour G et I : nécessite de construire un nouveau krigeage pour chaque point candidat
- Pour I : nécessite de calculer numériquement une intégrale pour chaque point candidat

On a diminué la dimension du problème, mais pas son coût

Planification adaptative en cherchant le maximum de variance

Algorithme

1. on construit un plan initial (taille k)
   Pour i allant de $k+1$ à n :
2. On cherche le point où la variance est maximum :
   $\mathbf{x}* = \arg\max_D s_K^2(\mathbf{x})$
3. On ajoute une nouvelle observation en ce point : $\mathbf{x}_i = \mathbf{x}*$
4. On répète les opérations 2 et 3

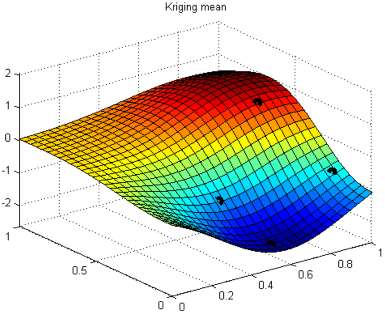
Avantages et inconvénients

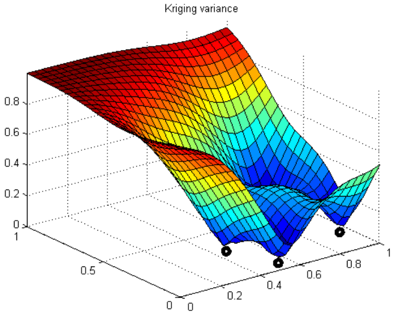
- "Coût" presque nul (pas d'inversion de matrice)
- Solution approchée du problème

Exemple

Plan de départ : 4 points

- IMSE = 0.5985
- maxMSE = 0.9991

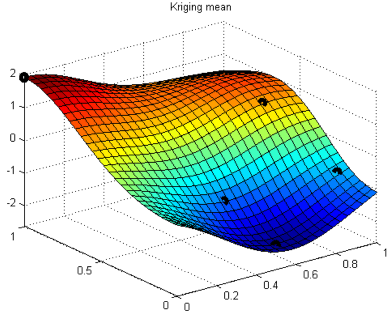


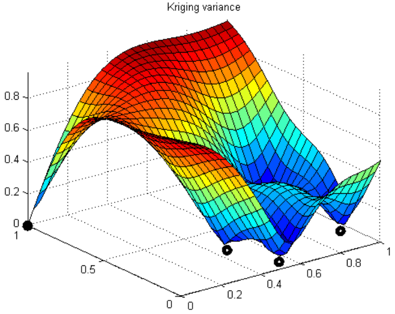


Exemple

5 points

- IMSE = 0.5462
- maxMSE = 0.9665

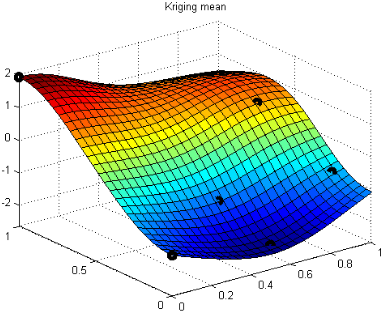


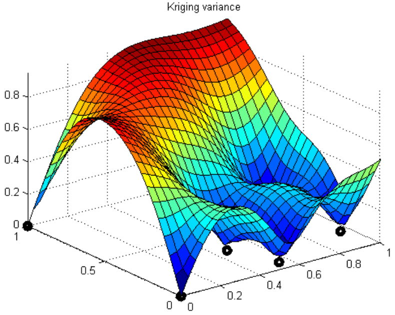


Exemple

6 points

- IMSE = 0.5011
- maxMSE = 0.9466

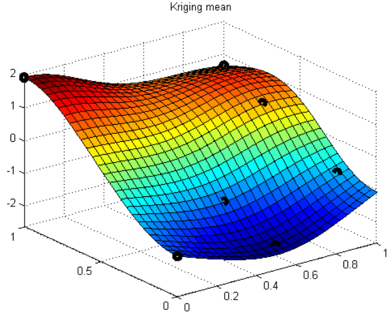


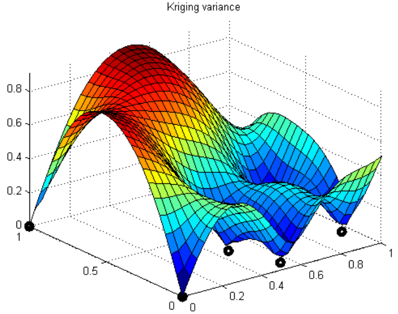


Exemple

7 points

- IMSE = 0.4619
- maxMSE = 0.9035

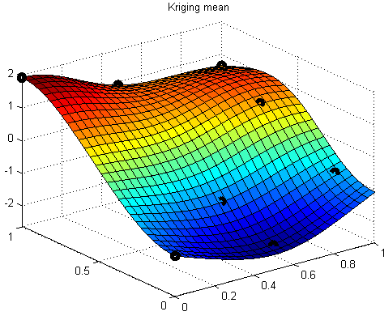


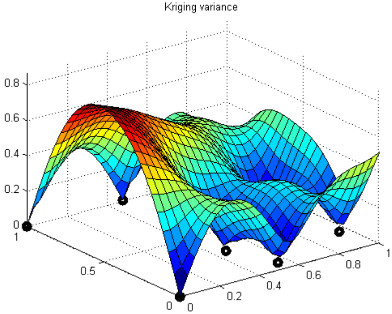


Exemple

8 points

- IMSE = 0.3885
- maxMSE = 0.8632

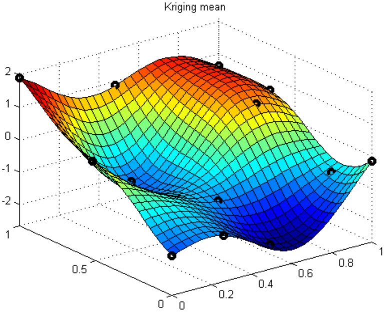


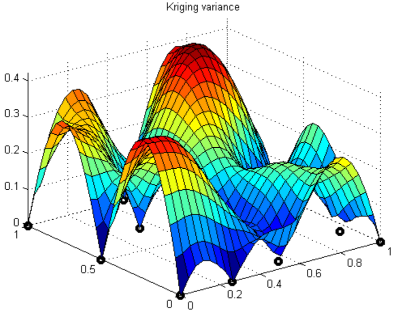


Exemple

14 points

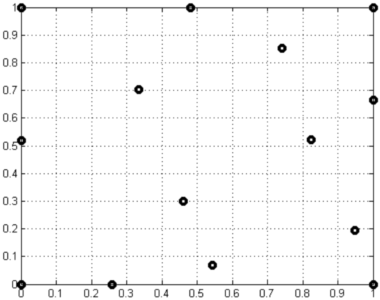
- IMSE = 0.2009
- maxMSE = 0.4226





Exemple : plan final

- Bon remplissage d'espace
- Tendance à échantilloner sur les bords



Remarques sur les plans adaptatifs

Plans adaptatifs et krigeage

Modèle "souple", donc bien adapté à la séquentialité.
Intérêt plus discutable en régression

Avantage des plans adaptatifs

Prise en compte la valeur des réponses !

Applications

- Optimisation $\Rightarrow$ EGO
- Planification ciblée

Conclusion (1/2)

Le modèle de krigeage

Plusieurs sources d'incertitude :

- Tendance
- Paramètres de covariance
- Variance de prédiction

Plans optimaux

- Pour l'estimation des paramètres : plus grande variation des situations
- Tendance seule : régression généralisée $\rightarrow$ D-optimalité
- Variance de prédiction : G-optimalité (maxMSE) / I-optimalité (IMSE)

Conclusion (2/2)

En pratique

- Plans space-filling
- Plans adaptatifs

Perspectives

- Plans adaptatifs : discipline ouverte
- Question ouverte : plans optimaux (notamment pour les observations bruitées)

Retour sur la définition

Objectifs

- Organisation réfléchie des expériences
- Génération contrôlée des données

pour **maximiser l'information obtenue**, soit :

- Mesurer l'influence des variables d'entrée sur la réponse ;
- Permettre l'utilisation de modèles reproduisant la complexité du processus étudié ;
- Maximiser la qualité de l'inférence du modèle

Objectif 1 : mesure des effets simples

Différents cas

Effets **linéaires** de chaque facteur sur la réponse :

- Plans en étoile
- Plans factoriels

Effets **linéaires et interactions** de chaque facteur sur la réponse :

- Plans factoriels

Si le nombre d'expériences possible est limité :

- Plans factoriels fractionnaires

En résumé : analyse de sensibilité

Cf. cours à suivre !

Objectif 2 : prédiction à l'aide d'un modèle (non présupposé)

Sans présupposition du modèle => remplissage d'espace

Différentes mesures de remplissage :

- Discrépance, erreur de quadrature => suites à faible discrépance
- Distortion, quadrature => tessellations centroïdales de Voronoï

Propriétés en projections

- Marginales d'ordre 1 : Hypercubes latins
- Niveau de stratification supérieur : hypercubes latins + tables orthogonales

Critères d'interdistances :

- distance minimale entre les points
- distance maximale entre un point du domaine et un point du plan

Objectif 3 : prédiction à l'aide d'un modèle de régression

Plans optimaux

Maximisation de l'information de Fisher => matrice de covariance des coefficients de la régression

- D-optimalité : déterminant
- A-optimalité, E-optimalité, etc.

Minimisation de l'erreur de prédiction :

- G-optimalité : variance de prédiction maximale

Théorème d'équivalence généralisé : les plans D- et G-optimaux sont les mêmes !

Pour aller plus loin

- Résultats théoriques : plans continus (proportions d'expériences)
- Compromis biais-variance : J-optimalité

Objectif 4 : prédiction à l'aide d'un modèle de krigeage

Apprentissage des paramètres : problème difficile !

- Tendance : D-optimalité avec résidus corrélés (cf. régression)
- Covariance : maximisation de la variété des interdistances

Prédiction à paramètres connus

Deux critères :

- I-optimalité : variance de prédiction moyenne
- G-optimalité : variance de prédiction maximum

Construction :

- Séquentielle ?
- En pratique : **remplissage d'espace**

Construction des plans

Problème d'optimisation globale, complexe !

Simplifications

- Problèmes discrétisés : LHS, grilles
- Algorithmes d'échange
- Planification séquentielle

Algorithmes

- D-optimalité : optimisation convexe
- LHS : recuit simulé
- Autres : méthodes globales, cf. cours et projet

Pour finir : le projet hélicoptère

Approche en trois étapes

1. Plan remplissant l'espace + construction d'un modèle
2. Enrichissement du plan + mise à jour du modèle
3. Optimisation à l'aide du modèle

Pourquoi autant d'étapes ?

- Pas de modèle présupposé => phase d'exploration nécessaire
- A modèle choisi : quelques points suffisent à beaucoup améliorer

Pourquoi ne pas optimiser directement après le plan initial ?

- Avec un modèle imprécis : beaucoup d'itérations nécessaires
- Enrichissement : ajout d'un ensemble d'observations (moins coûteux)
- Réduction du domaine pendant l'enrichissement ≈ pré-optimisation