Méthodes d'estimation

Marie-Odile Berger

November 8, 2011

Part I

Généralités sur l'estimation

Qu'est ce qu'un estimateur ?

Cadre expérimental: on considère un ensemble d'expériences donnant lieu à des mesures directes ou indirectes d'une grandeur θ que l'on veut estimer.

Problèmes:

- Trouver une estimation $\hat{\theta}$ de θ en fonctions des mesures
- Avec quelle précision est elle obtenue? ($\longleftrightarrow$ déterminer $var(\hat{\theta} - \theta)$).

Def: **statistique**
échantillon: vecteur composé de k v.a $x_1, ..., x_k$ indépendantes et de même distribution (ie k observations d'un même phénomène X). Toute fonction de $x_1, ..., x_k$ est appelé *statistique*.
Def: **estimateur**: fonction des résultats d'une épreuve (statistique) dont la réalisation est adoptée comme la valeur du paramètre inconnu θ.

Estimateur sans biais

Def: un estimateur $\hat{\theta}$ de θ est **sans biais** si $E(\hat{\theta}) = \theta$.

Exemples d'estimateurs

si les x_i sont les réalisations d'un v.a X, alors la moyenne est un estimateur sans biais de $E(X)$.
on a $E[x_i] = E[X]$

$$E(\tilde{x}) = E(\frac{1}{k}\sum_{i=1}^{i=k} x_i) = \frac{1}{k}\sum_{i=1}^{i=k} E(x_i) = E(X)$$

$s_x^2 = \frac{1}{k}\sum_{i=1}^{i=k}(x_i - \tilde{x})^2$ est un estimateur **biaisé** de $var(X)$.
un estimateur non biaisé de la variance est $\frac{1}{k-1}\sum_{i=1}^{k}(x_i - \tilde{x})^2$.

Estimation paramétrique/non paramétrique

Paramétrique: suppose que les données obéissent à un modèle (probabiliste ou non) défini par un nombre fini de paramètres à estimer. L'objectif est de déterminer les paramètres avec la meilleure précision possible.
Non paramétrique: tente de construire un modèle pour les données avec le moins possible d'hypothèses a priori.
exemple: l'estimation d'un loi de probabilité à partie d'un échantillon peut se faire à partir de la donnée a priori d'une loi (exemple gaussienne) ou sans hypothèse de loi par la méthode de Parzen.
En pratique, on s'intéressera essentiellement ici à l'estimation paramétrique.

Objectifs:

- Définir les modes d'estimation:
  moindre carrés, maximum de vraisemblance, méthodes des moments
- Calculer l'estimation, le mode étant choisi
  calcul explicite pour les moindres carrés, minimisation itérative le plus souvent
- Évaluer la qualité de l'estimation
  calculer la variance de l'estimation, estimer la qualité du *fitting*
- Estimer de façon robuste, c'est à dire sans tenir compte d'éventuelles données ne suivant pas le modèle.

De quoi dépendent les méthodes d'estimation

Les méthodes d'estimation sont très nombreuses. Elles dépendent

- de la connaissance de la loi physique théorique liant mesures et observations.
- de la connaissance de la loi de probabilité liant observation et mesures (modélisation probabiliste de l'erreur)
- de la connaissance d'une valeur approximative de l'estimation cherchée
- du critère retenu
  estimation d'un paramètre $\longleftrightarrow$ minimum d'une fonction de coût traduisant la compatibilité entre mesures et paramètres.

En général, le lien entre mesures et vecteur d'état est connu.

$$Z = f(X) + u$$

ou u est un bruit le plus souvent blanc dont la variance est connue et f souvent non linéaire.

Part II

Les méthodes d'estimation

Estimation au minimum de variance

On cherche un estimateur $\hat{\theta}$ de θ sans biais et de variance minimale

$$E(\hat{\theta}) = \theta, \quad Var(\hat{\theta}) \ minimale$$

Théorème: la meilleure estimation de X au minimum de variance est $E(X/Z)$.
cadre d'utilisation: les estimations récursives

le problème de la cible

Déterminer le centre de dispersion des tirs pour le faire coïncider avec le centre de la cible.
Pour améliorer l'efficacité, ne pas attendre que toutes les mesures soient faites mais apporter une correction après chaque tir $\rightarrow$ estimer les paramètres inconnus après chaque tir et adapter la correction apportée au paramètre en fonction de la mesure effectuée.
Voir le cours sur les estimateurs récursifs et le filtrage de Kalman.

Estimation au maximum de vraisemblance

Principe: Étant donné l'échantillon de mesures obtenues, la méthode du maximum de vraisemblance consiste à choisir parmi les valeurs possibles du paramètre, celle qui maximise la probabilité d'obtenir l'échantillon dont on dispose.
la loi $f(x|\theta)$ de la variable observée X est supposée connue.
exemple: l'observation suit une loi normale de paramètre $\theta = (m, \sigma)$: $f(x|\theta) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(\frac{x-m}{\sigma})^2}$.

soit u le vecteur des observations $[x_1, ..., x_n]^t$ résultat des épreuves $X_1, ..., X_n$.**On choisit $\hat{\theta}$ pour que la densité de probabilité des épreuves $X_1, ..., X_n$ prenne sa plus grande valeur pour les réalisations obtenues $x_1, ..., x_n$.**

$$\hat{\theta} = argmax \ \ g(x_1, ..., x_n|\theta)$$

$g(x_1, ..., x_n|\theta)$ est la fonction de vraisemblance.
$\hat{\theta}$ est l'estimation *au maximum de vraisemblance* du paramètre θ.
Remarque: mène souvent à des estimateurs biaisés

Méthode explicite de détermination de $\hat{\theta}$.

$$argmax g(x_1, ..., x_n|\theta) = argmax \ \ ln(g(x_1, ..., x_n|\theta))$$

Si les épreuves sont indépendantes:

$$g(x_1, ..., x_n|\theta) = \Pi_{k=1}^{k=n} f(x_k|\theta)$$
$$ln(g(x_1, ..., x_n|\theta)) = \Sigma ln(f(x_k|\theta))$$

si f est dérivable,

$$\sum \frac{\partial ln(f(x_k|\theta))}{\partial \theta} = 0$$

(si $\theta = (\theta_1, ..., \theta_r)$, il y a r équations scalaires).

Exemples d'estimation au maximum de vraisemblance

estimation du paramètre μ d'une distribution de Poisson.

$$P(x) = \frac{\mu^x e^{-\mu}}{x!}$$

estimer les paramètres λ et μ d'une distribution connaissant les observations $x_1, ..., x_n$ sachant que $P(x|\mu) = \frac{\mu^x e^{-\mu}}{x!}$ on a

$$\begin{aligned} 0 &= \sum \frac{\partial}{\partial \mu}(ln(\mu^{x_k} e^{-\mu}/x_k!) \\ &= \sum_k \frac{\partial}{\partial \mu}(x_k ln(\mu) - \mu) \\ &= \sum_k (x_k/\mu - 1) \end{aligned}$$

d'où $\mu = 1/n \sum_k x_k$

Exemples d'estimation au maximum de vraisemblance (2)

Détermination d'une densité de probabilité
Problème du calcul de $p(x|C_1)$ connaissant des éléments $x_1...x_n$ appartenant à C_1.
Si la forme générale de p est connue paramétriquement (ex loi de gauss, de poisson ...), alors p peut être déterminée au maximum de vraisemblance.
ex: on considère n mesures indépendantes $x_1...x_n$ suivant une loi de Maxwell de densité

$$p(x|A) = \sqrt{\frac{2}{\pi}} A^3 x^3 exp(-\frac{A^2 x^2}{2})$$

déterminer A au maximum de vraisemblance.

Exemple : apprentissage non supervisé [DH73]

Un exemple simple tiré de [DH73]

- on considère un ensemble de mesures pouvant provenir de 2 classes sans que l'on sache de laquelle.
- on connaît $p(\omega_1)$ et $p(\omega_2)$.
- on sait que $p(z|\omega_1)$ et $p(z|\omega_2)$ sont des gaussiennes de variance $\sigma^2 = 1$ mais de moyenne μ_1 et μ_2 inconnue.

Problème: estimer μ_1 et μ_2 à partir des observations.

$$\begin{aligned} p(x|\mu_1,\mu_2) &= p(x = X(\omega) \ et \ \omega \in \omega_1) + p(x = X(\omega) et \omega \in \omega_2) \\ &= p(x|\omega_1) * p(\omega_1) + p(x|\omega_2) * p(\omega_2) \\ &= \frac{1}{3}\frac{1}{\sqrt{2\pi}} exp(-\frac{1}{2}(x-\mu_1)^2) + \frac{2}{3}\frac{1}{\sqrt{2\pi}} exp(-\frac{1}{2}(x-\mu_2)^2) \end{aligned}$$

Apprentissage supervisé

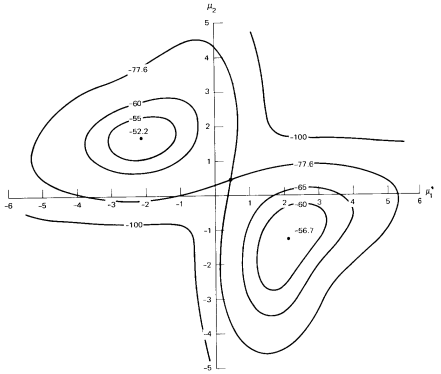
estimation au maximum de vraisemblance: maximiser $\sum log\ p(x|\mu_1,\mu_2)$.

TABLE 6-1. Twenty-five Samples from a Normal Mixture

k	x_k	(Class)
1	0.608	2
2	−1.590	1
3	0.235	2
4	3.949	2
5	−2.249	1
6	2.704	2
7	−2.473	1
8	0.672	2
9	0.262	2
10	1.072	2
11	−1.773	1
12	0.537	2

k	x_k	(Class)
13	3.240	2
14	2.400	2
15	−2.499	1
16	2.608	2
17	−3.458	1
18	0.257	2
19	2.569	2
20	1.415	2
21	1.410	2
22	−2.653	1
23	1.396	2
24	3.286	2
25	−0.712	1

Apprentissage non supervisé



2 maxima $(\mu_1,\mu_2) = (-2.130, 1.668)$ et $(\mu_1,\mu_2) = (2.085, -1.257)$.
En pratique, l'estimation au maximum de vraisemblance débouche sur des équations sans solutions directes et nécessite de recourir à des résolutions numériques.

Méthode des moments

Le cadre est le même que pour l'estimation au maximum de vraisemblance.
méthode: déterminer les paramètres en égalant les estimations empiriques et théoriques des moments.

exemple

Déterminer par la méthode des moments les estimations des paramètres m etα de la distribution uniforme sur $[m-\alpha, m+\alpha]$.
densité: $\frac{1}{2\alpha}1_{[m-\alpha,m+\alpha]}$
moment d'ordre 1 = m
moment d'ordre 2 $=\frac{3m^2+\alpha^2}{3}$
d'où

$$\begin{aligned} m &= \frac{1}{n}\sum x_k \\ \frac{3m^2+\alpha^2}{3} &= \frac{1}{n}\sum x_k^2 \end{aligned}$$

Méthode des moments

Remarque: La méthode des moments peut ne pas avoir de solutions si on prend davantage de moments (par exemple 3 moments ou plus pour 2 inconnues)
→ Utiliser une méthode généralisée en résolvant au mieux les équations des moments aux moindres carrés.
Exemple 2: Recaler deux formes en utilisant les moments. Trouver les paramètres de la transformation t.q les moments de la forme initiale transformée soient égaux aux moments de la forme cible.
Exemples d'articles utilisant des moments: [FCH05, Cha04]

Modèle déterministe et estimation aux moindres carrés

On se place dans le cas où le paramètre p et les mesures z sont décrites linéairement $A\theta = z$
Rappel: on peut choisir une norme significative du problème, par exemple $||Z||^2 = \sum \frac{1}{\sigma^2} z_i^2$ ou plus généralement $||Z||^2 = Z^t\Lambda^{-1}Z$ où Λ est la matrice de covariance des mesures.
On a l'estimation

$$\hat{\theta} = (A^t\Lambda^{-1}A)^{-1}A^t\Lambda^{-1}Z$$

Interprétation statistique des moindres carrés

Si on considère qu'en fait $Z = A\theta + v$ où v est un bruit gaussien de moyenne nulle et de variance $E(vv^t) = \Lambda$, alors

$$p(Z|\theta) = \frac{1}{(2\pi)^{n/2}\sqrt{det(\Lambda)}} exp(-\frac{1}{2}(Z-A\theta)^t\Lambda^{-1}(Z-A\theta))$$

puisque z est gaussienne d'espérance $A\theta$ et de variance Λ. $\hat{\theta}$ maximise $p(Z/\theta)$ → estimation au maximum de vraisemblance.

Exemple d'estimateur linéaire aux moindres carrés

Fusion de données Soient deux mesures (m_1, Λ_1) et (m_2, Λ_2) d'une même quantité. Déterminer une estimation m de cette quantité tenant compte de ces deux mesures.
Une façon de calculer m est de minimiser :

$$(m - m_1)^t \Lambda_1^{-1} (m - m_1) + (m - m_2)^t \Lambda_2^{-1} (m - m_2)$$

donc, en dérivant

$$\Lambda_1^{-1}(m - m_1) + \Lambda_2^{-1}(m - m_2) = 0$$

et

$$m = (\Lambda_1^{-1} + \Lambda_2^{-1})^{-1}(\Lambda_1^{-1} m_1 + \Lambda_2^{-1} m_2)$$

ou d'utiliser les moindres carrés avec $A = \begin{pmatrix} I \\ I \end{pmatrix}$

Exemple d'estimateur linéaire aux moindres carrés (2)

Modèles auto-régressifs Un processus temporel $\{y_t\}$ suivant un modèle auto-régressif est défini par

$$y_t + \sum_{k=1}^{p} a_k y_{t-k} = u_t$$

où u_t est un processus non corrélé de moyenne 0 et de variance σ_u^2.
modèle utilisé pour la parole et pour décrire des phénomènes temporels complexes mais assez réguliers.
Méthode 1: détermination des a_k aux moindres carrés
étant données des observations $y_0, ..., y_n$, on a

$$\underbrace{\begin{bmatrix} y_p \\ y_{p+1} \\ \dots \\ y_{n-1} \end{bmatrix}}_{y} + \underbrace{\begin{bmatrix} y_{p-1} & y_{p-2} & \dots & y_0 \\ y_p & y_{p-1} & \dots & y_1 \\ \vdots & \vdots & \vdots & \vdots \\ y_{n-2} & y_{n-3} & \dots & y_{n-p-1} \end{bmatrix}}_{Y} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = 0$$

solution: $a = -(Y^t Y)^{-1} Y^t y$

Modèles auto-régressifs: résolution statistique

méthode 2: utilisation des corrélations $r_k = cov(y_{t+k}, y_k)$

$$y_t + \sum_{k=1}^{p} a_k y_{t-k} = u_t$$

donc

$$\begin{aligned} cov(u_t, u_t) &= cov(u_t, y_t) \\ &= cov(y_t + \textstyle\sum_{k=1}^{p} a_k y_{t-k}, y_t) \\ &= r_0 + \textstyle\sum a_k r_k \end{aligned}$$

$$0 = cov(u_t, y_{t-1}) = cov(y_t + \sum_{k=1}^{p} a_k y_{t-k}, y_{t-1}) = r_1 + a_1 r_0 + a_2 r_1 + ... a_p r_{p-1}$$

d'où

$$\begin{bmatrix} r_0 & r_1 & \dots & r_p \\ r_1 & r_0 & \dots & r_{p-1} \\ \vdots & \vdots & \vdots & \vdots \\ r_p & r_{p-1} & \dots & r_0 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} \sigma_u^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (*)$$

Modèles auto-régressifs

Estimation de Yule-Walker: solution de (*) en remplaçant les covariances inconnues par les covariances estimées

$$\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = - \begin{bmatrix} \hat{r}_0 & \hat{r}_1 & \dots & \hat{r}_{p-1} \\ \hat{r}_1 & \hat{r}_0 & \dots & \hat{r}_{p-2} \\ \vdots & \vdots & \vdots & \vdots \\ \hat{r}_{p-1} & \hat{r}_{p-2} & \dots & \hat{r}_0 \end{bmatrix}^{-1} + \begin{bmatrix} \hat{r}_1 \\ \hat{r}_2 \\ \vdots \\ \hat{r}_p \end{bmatrix}$$

Moindres carrés non linéaires

L'approche aux moindres carrés se généralise à toute fonction liant mesure et paramètres $z_i = f_i(\theta)$.

$$\hat{\theta} = argmin \sum (z_i - f_i(\theta))^2$$

Dans ce cas, il n'y a pas (en général) de solution explicite et [$\hat{\theta}$ est obtenu par minimisation numérique.

Évaluer la précision de l'estimation

Le cas des moindres carrés L'erreur d'estimation dans un moindre carré est $(A^t \Lambda^{-1} A)^{-1}$.

Dans le cas non linéaire

$\hat{\theta} = argmin\phi(\theta) \quad \hat{\phi} = \phi(\hat{\theta})$

Il n'y a pas de raison de préférer θ à $\hat{\theta}$ lorsque

$$|\phi(\theta) - \hat{\phi}| \leq \epsilon$$

$\rightarrow$ définition d'une région d'indifférence dans l'espace des paramètres.

En utilisant un développement de Taylor à l'ordre 1

$$\phi(\theta) \approx \hat{\phi} + \hat{q}^t \delta\theta + \frac{1}{2} \delta\theta^t \hat{H} \delta\theta$$

où $\delta\theta = \theta - \hat{\theta}$, $\hat{q}$ est le gradient de ϕ calculé en $\theta = \hat{\theta}$ et $\hat{H}$ est le hessien de ϕ calculé en $\theta = \hat{\theta}$.

Évaluer la précision de l'estimation

$$\phi(\theta) \approx \hat{\phi} + \hat{q}^t \delta\theta + \frac{1}{2} \delta\theta^t \hat{H} \delta\theta$$

A l'optimum, $\hat{q} = 0$, donc

$$\phi(\theta) \approx \hat{\phi} + \frac{1}{2} \delta\theta^t \hat{H} \delta\theta$$

La région d'indifférence est donc définie par

$$\delta\theta^t \hat{H} \delta\theta \leq 2\epsilon$$

définit un ellipsoïde de dimension $dim(\theta)$.

Part III

Minimisation : considérations numériques

Pourquoi étudier les méthodes de minimisation numérique

[Cia82] est un bon ouvrage de référence.
pourquoi: les modes d'estimation au max de vraisemblance et les moindres carrés généralisés conduisent à minimiser des fonctions complexes → optimisation numérique
La difficulté varie selon que

- on cherche un optimum local (ou un optimum global avec une bonne solution initiale)
- on cherche un optimum global

Recherche d'un optimum local

Résolution de l'équation $f'(x) = 0$ on cherche à trouver le minimum d'une fonction en résolvant l'équation $f'(x) = 0$.
Méthode de Newton nécessite l'évaluation de f et f'.
on définit la suite

$$x_0 \quad x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$$

interprétation géométrique: x_{k+1} est l'intersection de la tangente à f en x_k avec l'axe des x.

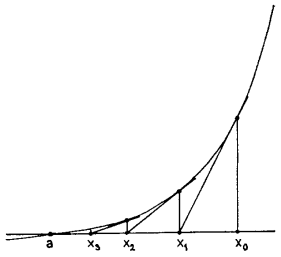


FIG. 7.5-1.

Recherche d'un optimum local: méthode de Newton

origine: si ζ est le zéro de f, le développement de Taylor de f au voisinage de x_0 est

$$f(\zeta) = 0 = f(x_0) + (\zeta - x_0)f'(x_0) + ...$$

en négligeant les termes d'ordre 2,

$$\zeta = x_0 - \frac{f(x_0)}{f'(x_0)}$$

nécessité de disposer d'une estimée initiale de la solution
généralisation en dimension supérieure
$f : R^n \to R$ Si la matrice de dérivée f' est une bijection en tout point x, on considère le schéma de Newton

$$x_{k+1} = x_k - \{f'(x_k)\}^{-1} f(x_k)$$

en pratique, calculer $f'(x_k)\}^{-1}$ à chaque itération est coûteux. → conserver la même matrice pendant p itérations.

$$x_{k+1} = x_k - \{f'(x_0)\}^{-1} f(x_k) \quad 0 \leq k \leq p-1$$
$$x_{k+1} = x_k - \{f'(x_p)\}^{-1} f(x_k) \quad p \leq k \leq 2p-1$$

minimisation: le cas mono-dimensionnel

Méthode de bracketing
utilisation de 3 points $a < b < c$

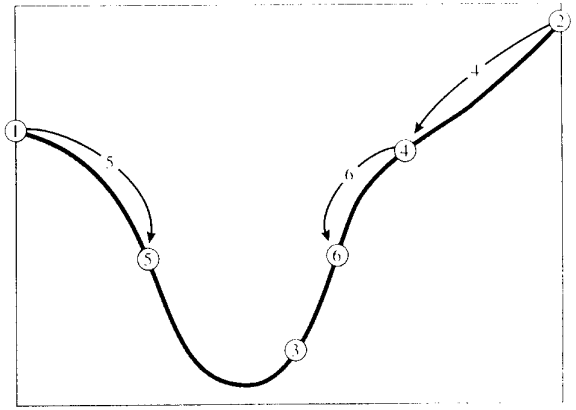


Figure 10.1.1. Successive bracketing of a minimum. The minimum is originally bracketed by points 1,3,2. The function is evaluated at 4, which replaces 2; then at 5, which replaces 1; then at 6, which replaces 4. The rule at each stage is to keep a center point that is lower than the two outside points. After the steps shown, the minimum is bracketed by points 5,3,6.

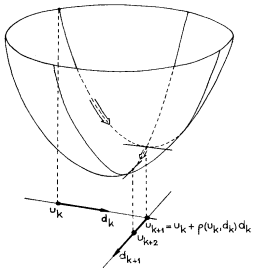
Le cas multi dimensionnel: les méthodes de relaxation

principe des méthodes de relaxation: se ramener à des minimisation successives de fonctions d'une seule variable réelle.

- solution initiale x_0.
- se donner une direction de descente d_k au point x_k.
- chercher le miminum de f le long de la droite passant par x_k et de vecteur directeur d_k. déterminer ρ tel que

$$f(x_k + \rho(x_k, d_k)d_k) = inf_\rho f(x_k + \rho d_k)$$

on pose $x_{k+1} = x_k + \rho(x_k, d_k)d_k$



Comment choisir les directions de descente?

- le plus simple: les axes des coordonnées pris de façon cyclique
- la direction du gradient (méthodes de gradient)
  pour éviter les minimisations coûteuses, méthodes de gradient à pas fixe:

$$x_{k+1} = x_k - \rho\nabla f(x_k)$$

  ρ étant à déterminer
- autres: ex powell

Recherche d'un optimum global et le recuit simulé

Méthode alternative: utiliser les méthodes d'optimisation locale en partant d'un grand nombre d'initialisations *bien réparties*... aléatoire
Principe du Recuit simulé [KGV83]: établir une analogie entre le problème à résoudre et un système physique évoluant à une certaine température.
Le recuit métallurgique:
pour trouver les états de basse énergie (états fondamentaux) d'un système complexe: réchauffer à une température élevée puis le refroidir (très) lentement $\rightarrow$ états cristallins de très faible énergie (forcer l'état du système dans des régions de basse énergie tout en lui évitant d'être piégé dans des états correspondants à des minima locaux d'énergie élevée).
Si la descente est trop rapide, matériau amorphe.

Principe du Recuit simulé

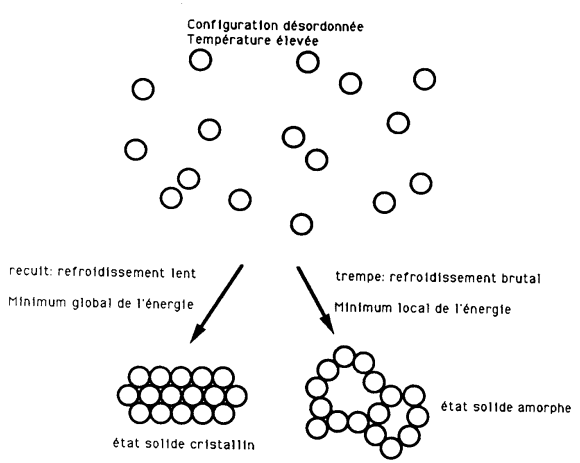


Figure 2.4: Comparaison des techniques du recuit et de la trempe.

les méthodes de descente locale conduisent à un refroidissement trop rapide.

Principe du Recuit simulé

- A partir d'une estimée initiale, un nouveau point est généré selon une fonction de distribution.
  Le point est accepté ou rejeté, selon une fonction d'acceptation h(x), qui est fonction d'un paramètre T appelé "température" , et de la différence des valeurs de la fonction à minimiser entre le point courant et le point précédent.
- Initialement, T est élevée, et le nouveau point est accepté près de la moitié du temps. Quand la température baisse, il devient de plus en plus improbable d'accepter un point dont la valeur est plus grande que la valeur du point précédent, ce qui effectue bien une minimisation de la fonction.
- Accepter un point même s'il donne une valeur de la fonction plus grande que la valeur précédente permet à l'algorithme de sauter les minima locaux et de trouver le minimum global.

Mise en œuvre du recuit simulé

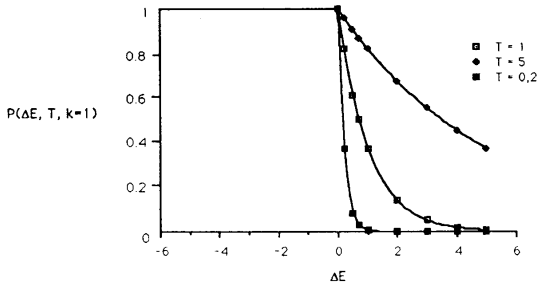


Figure 2.6: Probabilités d'acceptation d'une transformation élémentaire en fonction de ΔE.

Mise en œuvre du recuit simulé

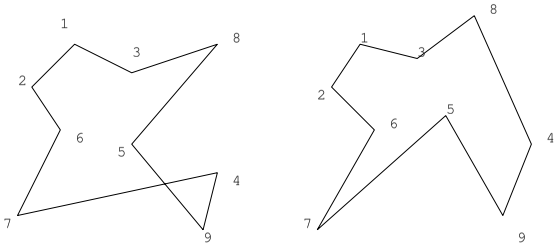
[Hér91, KGV83, MRTT53, 2] **Thermalisation à une température T** algorithme de type *monte carlo* engendrant une séquence d'états convergeant vers une distribution de Boltzman.

1. soit une configuration initiale (aléatoire) $x = (x_1, ..., x_n)$
2. à partir de cet état, effectuer un petit déplacement (par exemple en changeant aléatoirement une des coordonnées). Soit x' le nouvel état.
3. évaluation du différentiel d'énergie $\Delta E = E_x - E_{x'}$
4. si $\Delta E < 0$, accepter le déplacement ($\rightarrow$ on a un meilleur minimum)
5. si $\Delta E \geq 0$, accepter le déplacement avec la probabilité $exp(-\Delta E/kt)$. (en pratique, tirer un nombre aléatoire dans [0,1] et le comparer à $exp(-\Delta E/kt)$. s'il est inférieur on accepte la transformation, sinon elle est refusée.
6. répéter les étapes 2 à 5 jusqu'à atteinte de l'équilibre

Exemple : Le problème du voyageur de commerce [PFTV88]

Étant données n villes déterminer le plus court trajet passant par toutes ces villes exactement une fois.
combinatoire élevée: $(n-1)!/2$.

- état: une permutation de $1..n$ interprétée comme l'ordre dans lequel les villes sont visitées
- changement aléatoire de configuration: tirer 2 villes au hasard et inversion le trajet sur ce tronçon

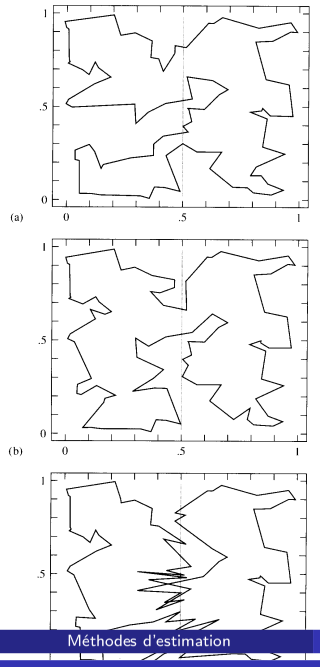


1 2 6 7 4 9 5 8 3 1 2 6 7 5 9 4 8 3

Exemple : Le problème du voyageur de commerce

- fonction a minimiser $\Sigma\sqrt{(x_i - x_{i+1})^2 + (y_i - y_{i+1})^2}$
- mise en œuvre: T est maintenue tant que on n'a pas atteint $100 \times n$ reconfigurations ou si on a atteint $10 \times n$ reconfigurations acceptées

Recuit simulé



Bibliographie

F. Chaumette.
Image moments: a general and useful set of features for visual servoing.
IEEE Trans. on Robotics, 20:713–723, 2004.

P. G. Ciarlet.
Introduction à l'analyse numérique matricielle et à l'optimisation.
Masson, 1982.

R. O. Duda and P. E. Hart.
Pattern Classification and Scene Analysis.
Wiley-InterScience, 1973.

R. Fletcher
Practical Methods of Optimization, Wiley, 1987

Bibliographie

A. Foulonneau, P. Charbonnier, and F. Heitz.
Affine-Invariant Geometric Shape Priors. for Region-Based Active Contours.
Rapport de recherche RR-AF01-2005, LSIIT, 2005.

S. Geman and D. Geman.
Stochastic Relaxation, Gibbs Distribution, and Bayesian Restoration of Images.
IEEE Transactions on PAMI, 6:721–741, 1984.

L. Hérault.
Reseaux de neurones récursifs pour l'optimisation combinatoire ; application à la théorie des graphes et à la vision par ordinateur.
Thèse de doctorat, Institut National Polytechnique de Grenoble, February 1991.

Bibliographie

- S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi.
  Optimization by Simulated Annealing.
  Science, 220:671–680, 1983.
- N. Metropolis, A. W. Rosenbluth, A. H. Teller, and E. Teller.
  Equations of state Calculations by Fast Computing Machines.
  J. Chem. Phys, 21:1087–1092, 1953.
- W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling.
  Numerical Recipes in C, The Art of Scientific Computing.
  Cambridge University Press, 1988.
- V. Pugachev.
  Théorie des probabiliteés et statistique mathématique.
  Editions de Moscou, Mir, 1982.

Bibliographie

- J. Stoer and R. Bulirsch.
  Introduction to Numerical Analysis.
  Springer-Verlag, New York, 1980.
- P. van Laarhoven and E. H. L. Aarts.
  Simulated Annealing: theory and Applications.
  Mathematics and its Applications. D. Reidel Publishing Company, Dordrecht, Holland, 1987.