

Introduction à l'estimation Robuste

Marie-Odile Berger

Qu'est ce que l'estimation robuste?

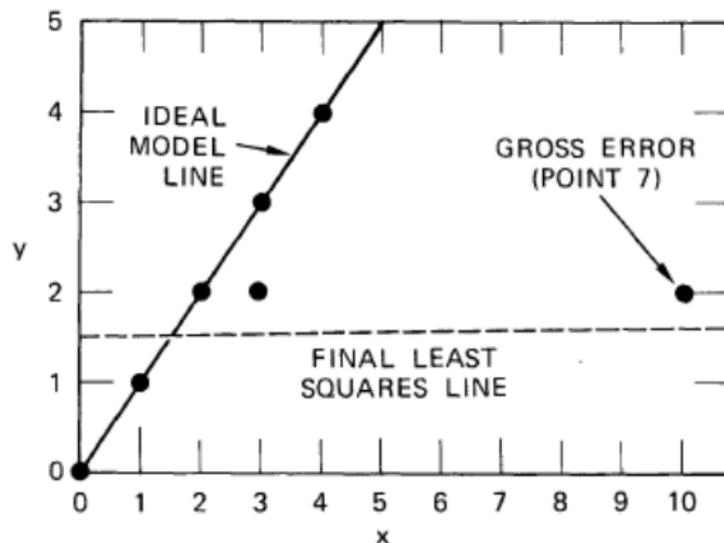
estimateur robuste = estimateur insensible à de *petites* déviations vis à vis du modèle pour lequel l'estimateur a été optimisé.

Que signifie *petites déviations*?

- toutes les données sont entachées d'une petite erreur
- quelques données sont entachées d'une très grosse erreur (**outlier**).

Un exemple frappant (Fishler & Bolles)

PROBLEM: Given the set of seven (x,y) pairs shown in the plot, find a best fit line, assuming that no valid datum deviates from this line by more than 0.8 units.



POINT	x	y
1	0	0
2	1	1
3	2	2
4	3	2
5	3	3
6	4	4
7	10	2

Un exemple frappant (Fishler & Bolles)

COMMENT: Six of the seven points are valid data and can be fit by the solid line. Using Least Squares (and the "throwing out the worst residual" heuristic), we terminate after four iterations with four remaining points, including the gross error at (10,2) fit by the dashed line.

SUCCESSIVE LEAST SQUARES APPROXIMATIONS		
ITERATION	DATA SET	FITTING LINE
1	1, 2, 3, 4, 5, 6, 7	$1.48 + .16x$
2	1, 2, 3, 4, 5, 7	$1.25 + .13x$
3	1, 2, 3, 4, 7	$0.96 + .14x$
4	2, 3, 4, 7	$1.51 + .06x$

COMPUTATION OF RESIDUALS				
POINT	ITERATION 1 RESIDUALS	ITERATION 2 RESIDUALS	ITERATION 3 RESIDUALS	ITERATION 4 RESIDUALS
1	-1.48	-1.25	-.96*	—
2	-0.64	-0.38	-.10	-.57
3	-0.20	0.49	.76	.37
4	0.05	0.36	.63	.31
5	1.05	1.36*	—	—
6	1.89*	—	—	—
7	-1.06	-0.57	-.33	-.11

- l'hypothèse d'erreur gaussienne est souvent sous entendue dans les modèles d'estimation mais pas vérifiée en réalité.
- les erreurs aberrantes sont mal prises en compte dans l'estimation classique.

Exemple: les moindres carrés sont très sensibles à la présence de données aberrantes, même en très petit nombre.

Remarque: les méthodes dépendent de la facilité d'identifier les mesures erronées (connaissances a priori).

Ce que l'on attend d'un estimateur robuste [zhang00]

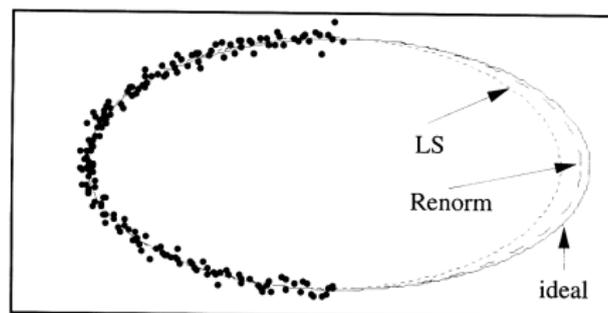


Fig. 5: Comparison between Linear Least-Squares (*LS*, in dotted lines) and Renormalization technique (*Renorm*, in dashed lines)

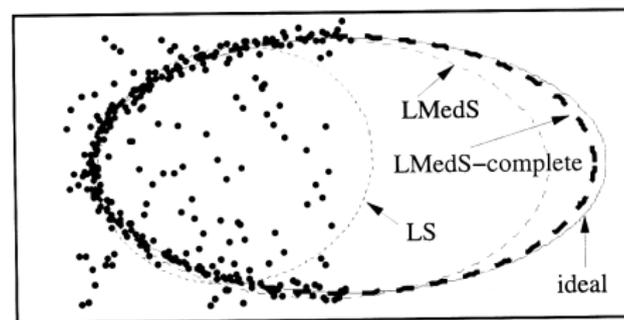


Fig. 7: Comparison of different techniques in the presence of outliers in data

Pourquoi l'estimation aux moindres carrés n'est pas robuste

Les Moindres carrés cherchent à minimiser un résidu quadratique:

$$\sum r_i^2(p)$$

En dérivant

$$\frac{\partial}{\partial p} \sum r_i^2(p) = 0$$

$$\sum r_i(p) \frac{\partial r_i(p)}{\partial p_j} = 0 \text{ pour } j = 1..r$$

$r_i(p)$ est élevé pour les outliers \rightarrow ces termes sont davantage pris en compte dans l'estimation.

- La transformée de Hough (classification)
- Diagnostic des données aberrantes
- M estimateur: modification de la fonction d'erreur de façon à ce que les mesures de résidu élevé soient peu prises en compte.
- Moindres carrés médians: utilisation d'un estimateur robuste *la médiane*
- utilisation d'ensemble de consensus: prélever dans les données des ensembles de données cohérentes.

Approche globale:

Etant donnée une quantification de l'espace de paramètres (cellules), déterminer pour chaque cellule l'ensemble des mesures compatibles avec la valeur du vecteur d'état correspondant à la cellule.

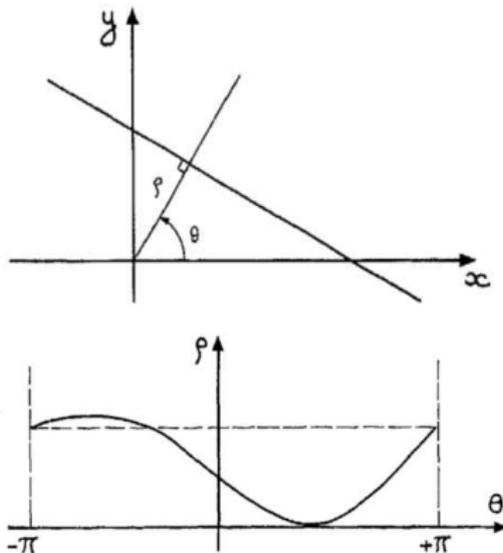
Les points d'accumulation indiquent les ensembles cohérents de mesure et la valeur correspondante du paramètre.

Difficultés:

- pouvoir borner l'ensemble des paramètres + découpage en cellules de taille équivalente
- taille de l'ensemble des cellules (estimation fine \rightarrow nombre gigantesque de cellules).
- les point d'accumulation sont souvent des taches très étalées.

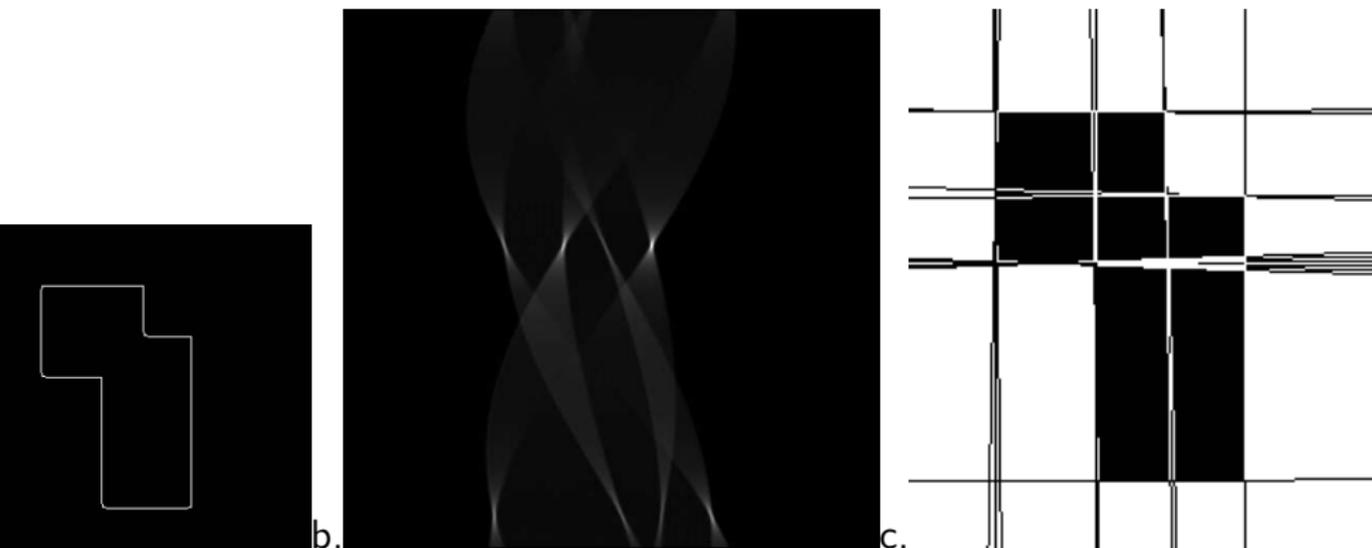
La transformée de Hough

Pour une représentation en coordonnées polaires $\rho = x\cos(\theta) + y\sin(\theta)$



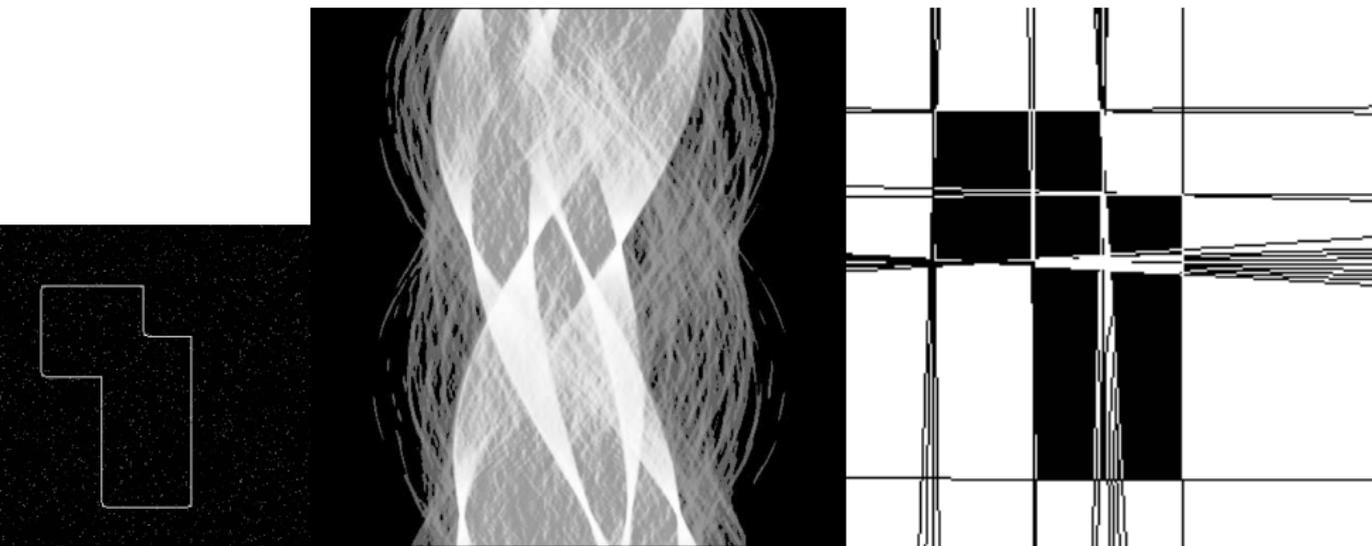
un point (x_i, y_i) donne lieu dans l'espace des paramètres à une courbe $\rho = x_i\cos(\theta) + y_i\sin(\theta)$.

La transformée de Hough: exemple



a: l'image initiale; b: l'accumulateur, c: résultat

La transformée de Hough: exemple bruité



C'est une mauvaise idée en général en présence d'outliers!

Démarche

- 1 Détermination d'une estimée initiale de p
- 2 Calcul du résidu correspondant pour chaque mesure
- 3 Élimination des données présentant un fort résidu (seuil)
- 4 Mise à jour de l'estimation de p en utilisant les données jugées cohérentes Puis retour à 2

Méthode dangereuse en présence d'outliers!

résidu: écart entre le modèle et les mesures disponibles.

cas de la régression par une droite $y = ax + b$. on considère les mesures y_i obtenues pour des abscisses x_i données. on $\hat{y}_i = ax_i + b$.

$$r_i^2 = (y_i - ax_i - b)^2$$

$\sum r_i^2$ mesure l'adéquation de l'ensemble des données au modèle.

Rq: on peut utiliser d'autres formes de résidus $\sum |y_i - \hat{y}_i|$, $\sum (y_i - \hat{y}_i)^4 \dots$ lié au modèle de l'erreur toléré pour les données.

Évolution des idées:

- remplacer la minimisation de la somme des carrés des résidus $\sum r_i^2$ par une fonction des résidus $\sum \rho(r_i)$. \rightarrow notion de M estimateurs
- remplacer la somme par la médiane qui est un estimateur robuste \rightarrow Least Median Square estimator (LMS)

Part I

Les M estimateurs

$$\text{minimiser } \sum \rho(r_i)$$

où ρ est une fonction positive, symétrique, avec un minimum en 0.

$$\frac{\partial}{\partial p} \sum \rho(r_i) = 0$$

$$\sum \psi(r_i) \frac{\partial r_i}{\partial p_j} = 0 \text{ pour } j = 1..r$$

où $\psi(x) = \frac{d\rho(x)}{dx}$ (**fonction d'influence**)

en posant $w(x) = \frac{\psi(x)}{x}$, on a

$$\sum w(r_i) r_i \frac{\partial r}{\partial p_j} = 0 \text{ pour } j = 1..r \quad (1)$$

qui correspond à la minimisation d'un problème de moindres carrés pondérés

$$\min \sum w_k(r_i) r_i^2$$

Quelques M estimateurs classiques:

- $\rho(x) = x^2$: estimateur non robuste car la fonction d'influence n'est pas bornée
- $\rho(x) = |x|$, fonction d'influence constante
- Fonction de Huber:

$$\rho(x) = \begin{cases} \text{si } |x| < k, \rho(x) = x^2/2 \\ \text{si } |x| > k, \rho(x) = k(|x| - k/2) \end{cases}$$

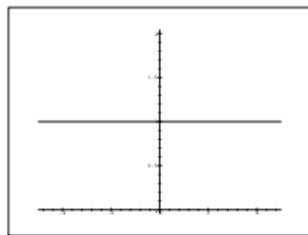
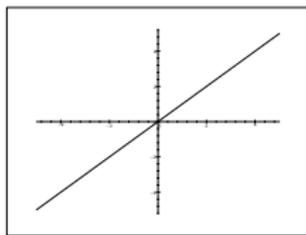
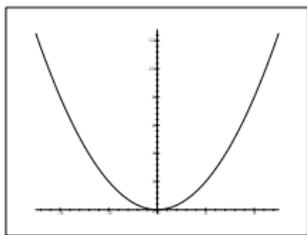
- fonction de Cauchy (ou de Lorenz)

$$\rho(x) = c^2/2 \log(1 + (x/c)^2)$$

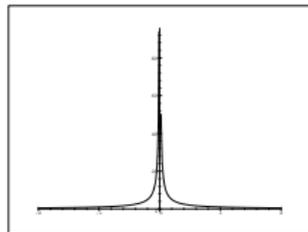
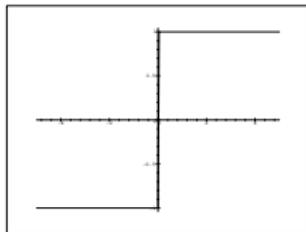
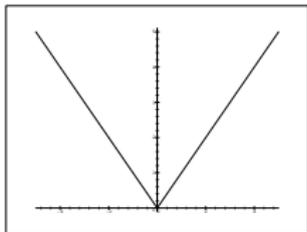
$$\psi(x) = \frac{x}{1 + \frac{1}{2}(x/c)^2}$$

Quelques M estimateurs classiques:

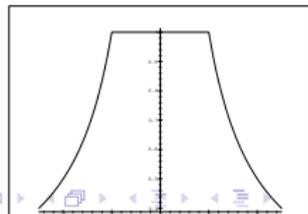
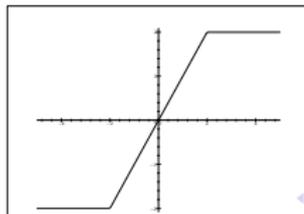
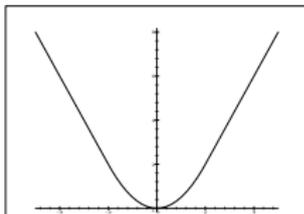
L_2



L_1



Huber



Quelques M estimateurs classiques:

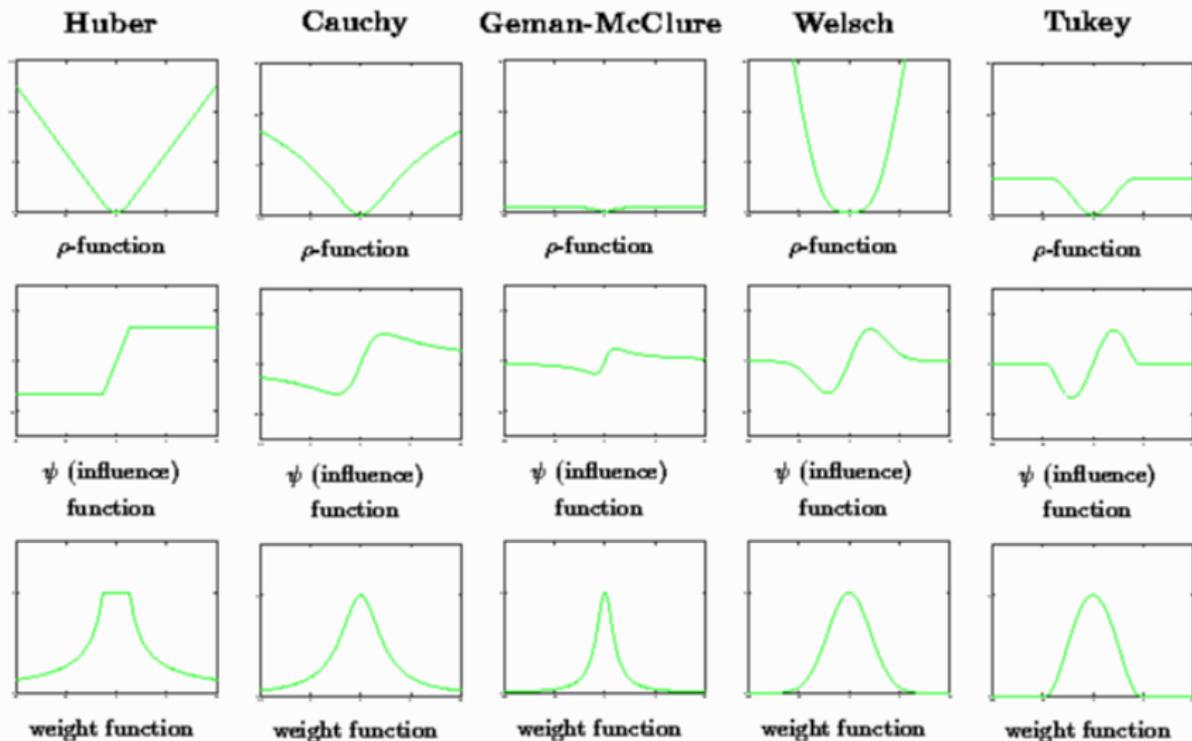
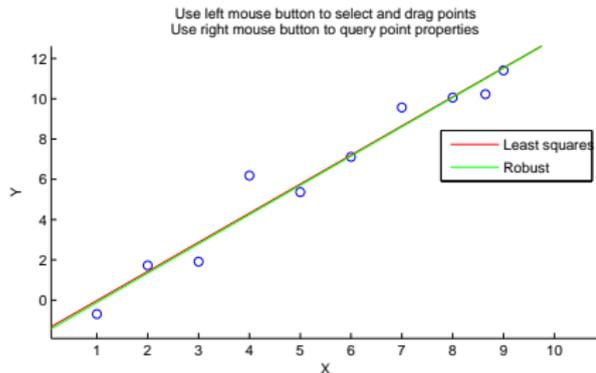


Figure 4: Graphic representations of a few common M-estimators

Remarque: pour utiliser les M-estimateurs, il faut

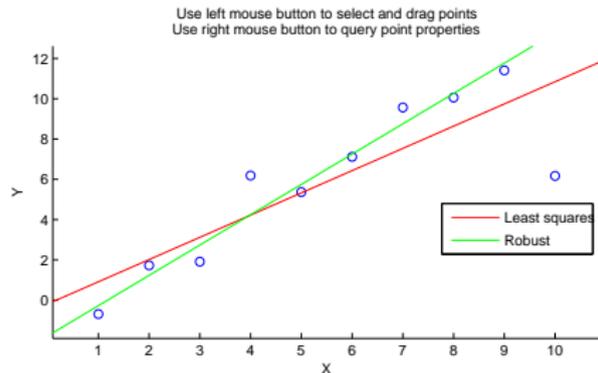
- choisir une distribution d'erreur
- résoudre l'équation (1), en général non linéaire → estimation numérique

Exemple d'utilisation des M estimateurs



Least squares: $Y = -1.46425 + 1.44491 \cdot X$
Robust: $Y = -1.56432 + 1.45339 \cdot X$

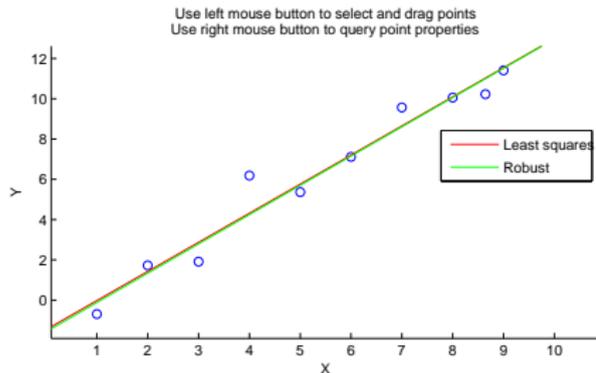
RMS error = 0.908837
RMS error = 0.924969



Least squares: $Y = -0.188327 + 1.10351 \cdot X$
Robust: $Y = -1.77278 + 1.50415 \cdot X$

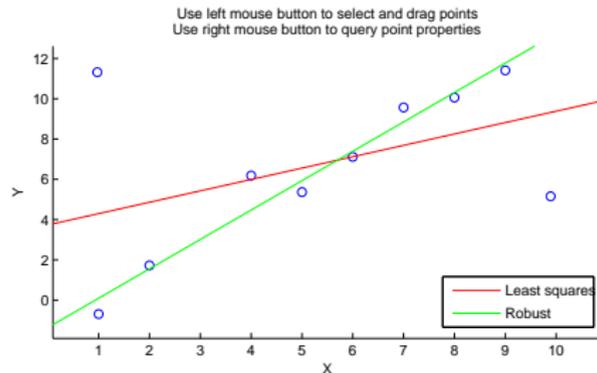
RMS error = 2.21375
RMS error = 1.43663

Exemple d'utilisation des M estimateurs



Least squares: $Y = -1.46425 + 1.44491 \cdot X$
Robust: $Y = -1.56432 + 1.45339 \cdot X$

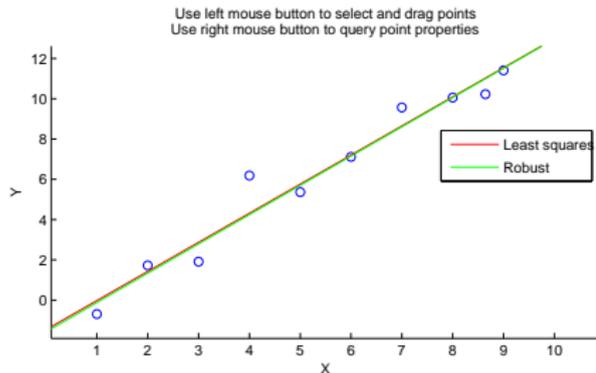
RMS error = 0.908837
RMS error = 0.924969



Least squares: $Y = 3.73471 + 0.564821 \cdot X$
Robust: $Y = -1.36121 + 1.45915 \cdot X$

RMS error = 3.81546
RMS error = 2.21661

Exemple d'utilisation des M estimateurs

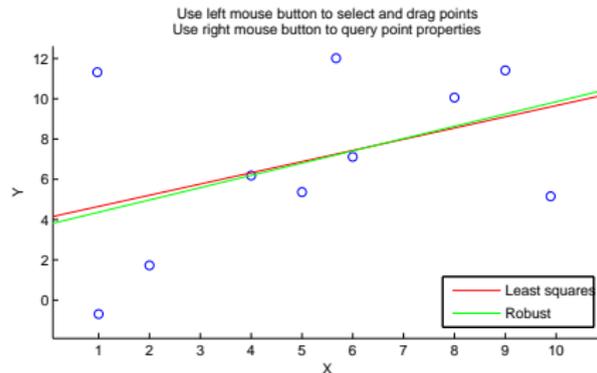


Least squares: $Y = -1.46425 + 1.44491 \cdot X$

RMS error = 0.908837

Robust: $Y = -1.56432 + 1.45339 \cdot X$

RMS error = 0.924969



Least squares: $Y = 4.10134 + 0.55564 \cdot X$

RMS error = 4.1477

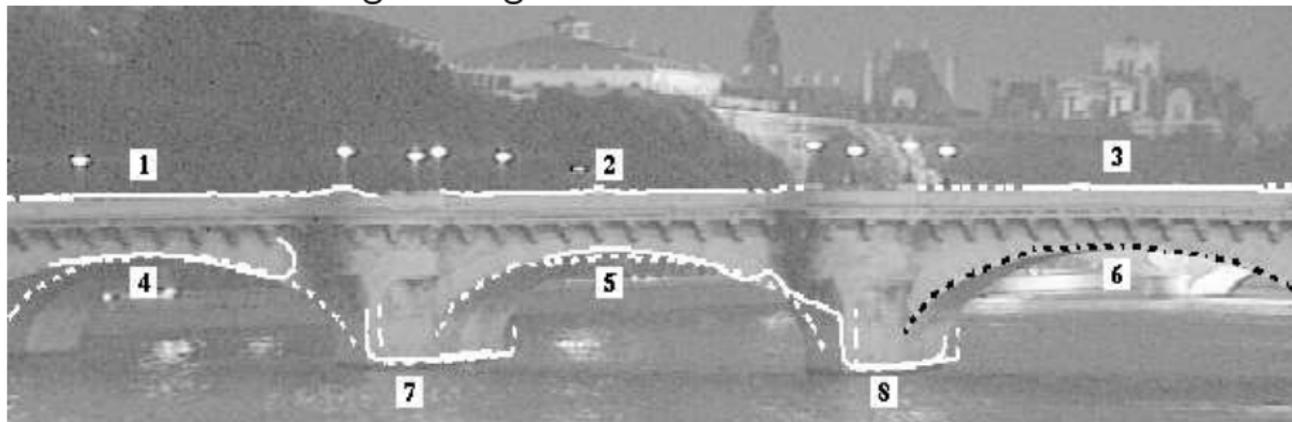
Robust: $Y = 3.75765 + 0.609081 \cdot X$

RMS error = 4.54283

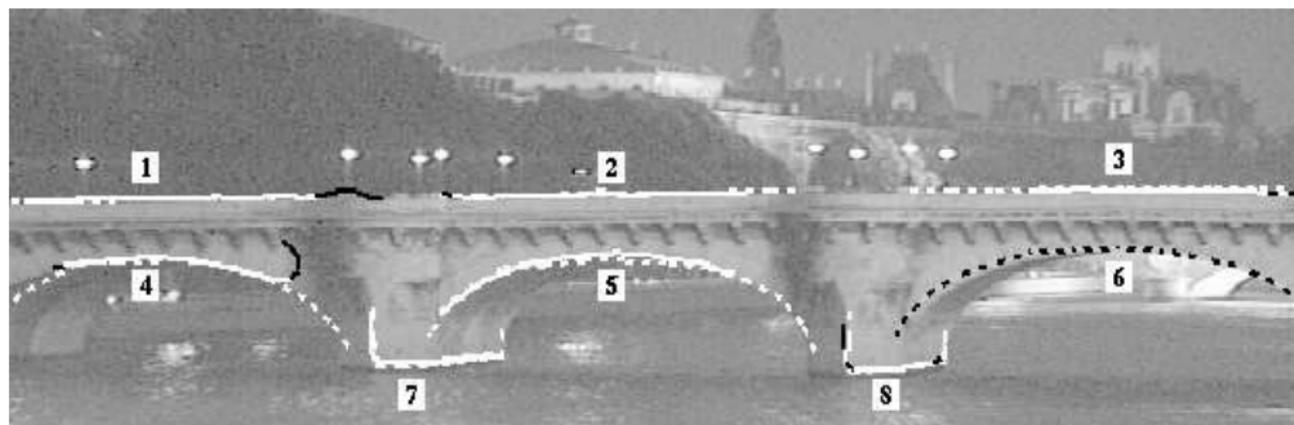
Exemple d'utilisation des M estimateurs

But: calculer le point de vue d'une caméra à partir d'un modèle et de données images.

caractéristique: les données images sont obtenues par suivi et peuvent être entachées d'erreur légère ou grossière.



Exemple d'utilisation des M estimateurs



en noir, parties éliminées par l'estimateur robuste



Résultat final



Part II

changer d'estimateur: moindres carrés médians

Un exemple d'estimateur robuste: la médiane

Espérance mathématique (moyenne)

$$E[X] = \int xf(x)dx \text{ ou } \sum x_i \mathcal{P}(X = x_i)$$

où f est la densité de probabilité de la variable aléatoire X . **Médiane:**

$$\int_{-\infty}^{\text{mediane}} f(t)dt = \int_{\text{mediane}}^{\infty} f(t)dt = \frac{1}{2}$$

pour des mesures discrètes et équiprobables:

$$\begin{cases} x_{med} = x_{\frac{N+1}{2}} \text{ si } N \text{ est impair} \\ x_{med} = \frac{1}{2}(x_{N/2} + x_{N/2+1}) \text{ si } N \text{ est pair} \end{cases}$$

La médiane est un estimateur **robuste**, cad peu sensible au bruit.

Estimateur robuste

pour un paramètre p donné, on peut calculer *mediane* r_i^2 . Parmi tous les p possibles, choisir celui qui minimise la médiane

estimateur : $\text{Min}_p \text{ med}_i r_i^2$

On explore seulement un ensemble de solutions plausibles:

- 1 Génération de m sous ensembles $(\mathcal{F}_1, \dots, \mathcal{F}_m)$ à q éléments (tirage aléatoire)
- 2 pour chaque $\mathcal{F}_{i_{1 \leq i \leq m}}$, calculer la valeur correspondante p_i du paramètre.
- 3 pour chaque $\mathcal{F}_{i_{1 \leq i \leq m}}$, calculer le résidu correspondant au paramètre p_i sur l'ensemble des données. calculer sa médiane.
- 4 $\hat{p} =$ valeur p_i minimisant la médiane $median_{j=1..n} r_j^2(p_i)$

Comment choisir m ?

[Zha 95, Rou 87]

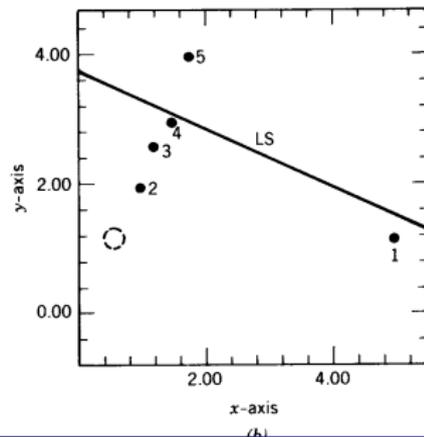
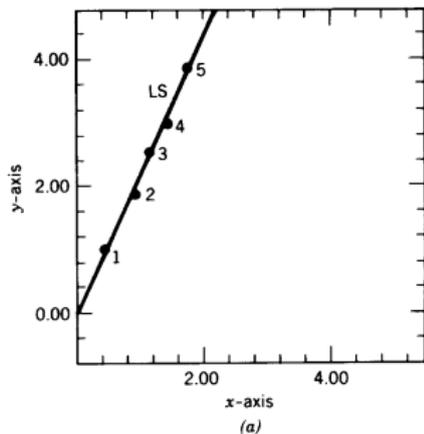
si la fraction ϵ d'outliers est connue, choisir m pour que la probabilité **qu'au moins** un des m sous ensembles soit correcte (ie ne contienne pas d'outliers) est proche de 1.

$$\text{Proba}(\text{au moins un des ensembles est correcte}) = 1 - [1 - (1 - \epsilon)^q]^m$$

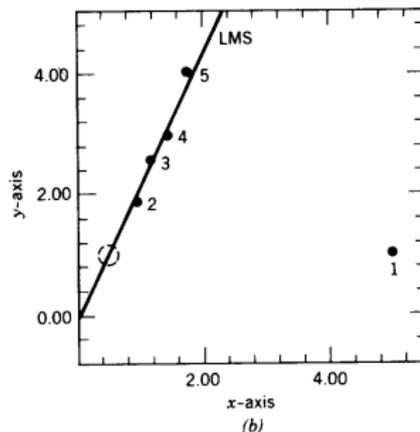
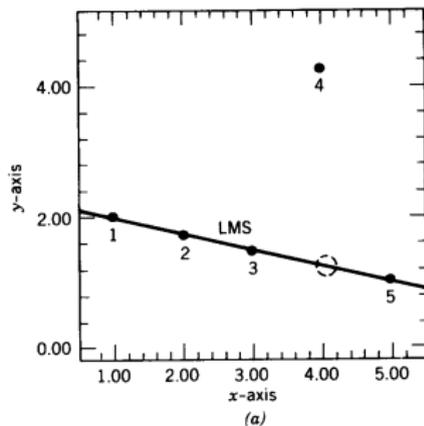
si $\epsilon = 40\%$, $q=8$, pour $P=.99$

$$m = \frac{\ln(.001)}{\ln(1 - .6^8)} \approx 272$$

Résultats: LS classique



Résultats: Moindres carrés médians



Les algorithmes de type *Random Sample Consensus* (RANSAC)[Fis 80]

Méthode: plutôt que d'utiliser tous les points disponibles et tenter d'éliminer les mauvais après, utiliser un petit nombre de points pour calculer une solution initiale et déterminer l'ensemble des points compatibles avec cette estimation.

Algorithme RANSAC

on considère un ensemble de données D . Soit q le nombre minimal de données permettant d'estimer le paramètre.

choisir aléatoirement un ensemble S_1 de q points dans D et calculer le paramètre p .

Utiliser p pour déterminer le sous ensemble de point S_1^* compatible avec p à une tolérance près. (S_1^* est l'ensemble de *Consensus*).

- si $\#(S_1^*) > \text{seuil}$, utiliser S_1^* pour calculer une estimation (aux moindres carrés) prenant en compte tous les points de l'ensemble de consensus.
- si $\#(S_1^*) \leq \text{seuil}$ choisir un nouvel ensemble aléatoire et répéter le processus.

questions

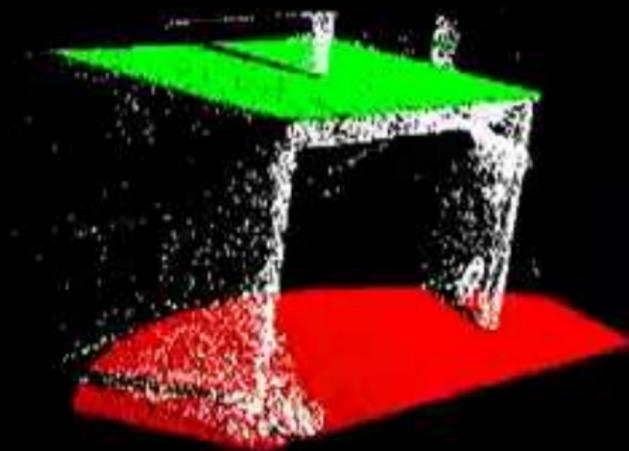
- Comment déterminer la tolérance permettant de former l'ensemble de consensus (problème de l'échelle, [Wang 04])
- Choix du seuil d'arrêt de la méthode (taille souhaitée de l'ensemble de consensus)

Exemple: extraction de plans dans un nuage de points

Extracting sets of indices

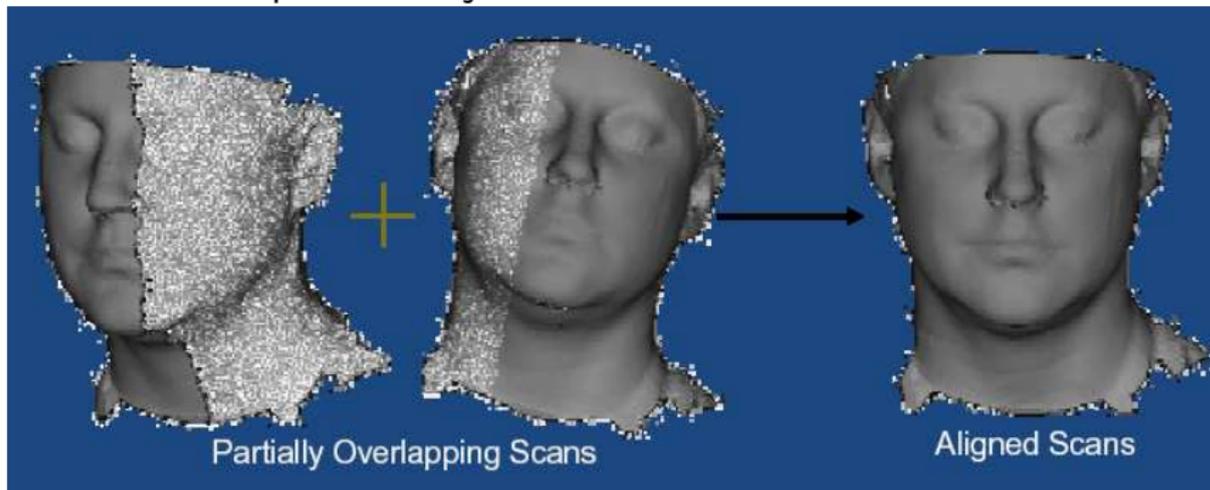
Partager

Plus d'infos



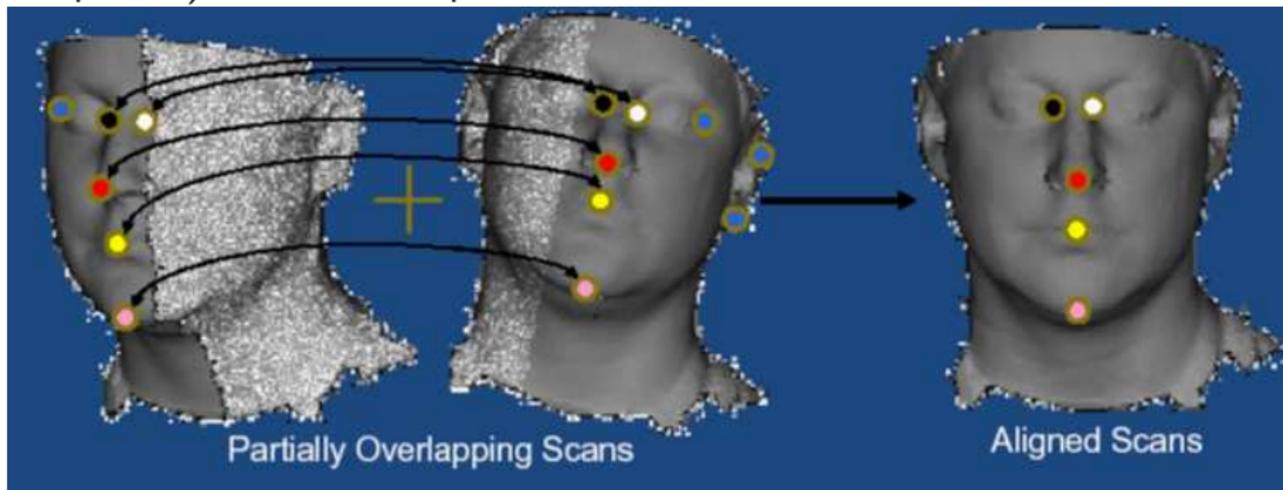
Exemple: recollement et alignement d'objets

Étant donnés deux scans partiellement recouvrant, les recoller pour avoir un modèle complet de l'objet.



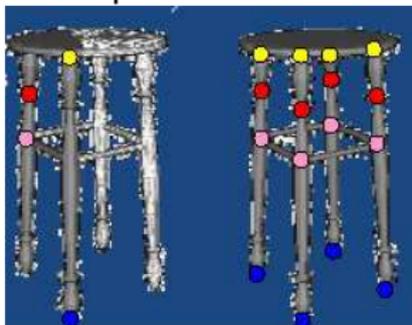
Exemple: recollement et alignement d'objets

problème: identifier les parties recouvrantes (ou des correspondances dans ces parties) et s'en servir pour fusionner



Difficultés de la mise en correspondance

Même avec de bons descripteurs(ici basés sur la courbure), il y a souvent des ambiguïtés de mise e correspondance en raison des symétries:



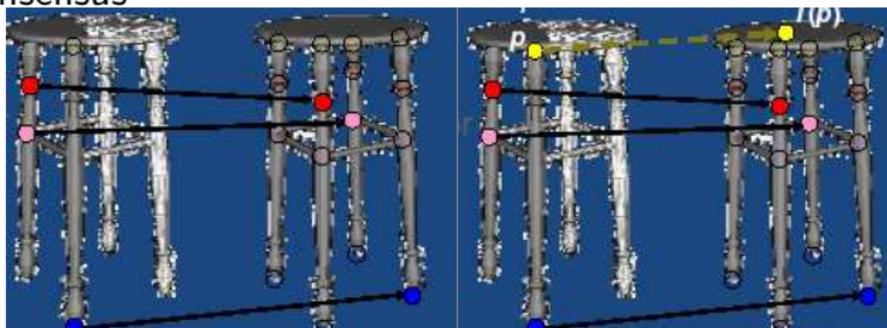
Utiliser RANSAC pour émettre des hypothèses de mise en correspondance et trouver la meilleure transformation (3 points en correspondance suffisent pour calculer la correspondance)

- choisir aléatoirement 3 points sur sur la surface source
- Pour chaque ensemble de 3 points sur la cible (compatibles!!)
 - Calculer la transformation T telle que $T(M_i) = P_i, i = 1..3$
 - Calculer $distance(T(p), cible)$ pour tous le autres points et en déduire l'ensemble de consensus

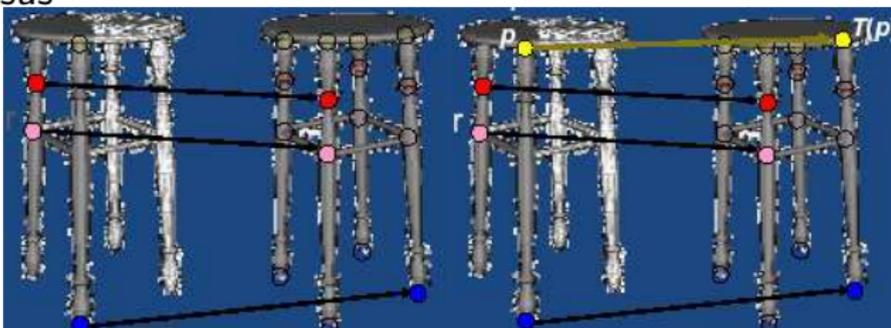
Choisir la transformation T ayant le plus grand ensemble de consensus

Utilisation de RANSAC

mauvais consensus



bon consensus



Construction de cartes de grands environnements en collant des cartes plus petites acquises par un observateur en mouvement (ex Kinect)

Améliorations pour être plus rapide, plus robuste, plus précis... [Choi 09]

- prendre en compte pas seulement le nombre de données à une distance $< s$ mais la somme des distances \rightarrow favoriser la précision
- procéder à un échantillonnage guidé plutôt qu'un échantillonnage aléatoire pour aller plus vite (en fonction de l'application, en abandonnant une hypothèse dès que certaines données sont à une distance $> s$).
- abandonner une hypothèse sur la base d'une évaluation partielle
- Évaluation adaptative du paramètre de distance au modèle et du nombre d'itérations
- permettre l'estimation de plusieurs modèles simultanément [Fusiello2008]

-  P. Besl and N. McKay.
A method for registration of 3-d shapes.
IEEE Transactions on PAMI, 14(2):239–256, February 1992.
-  S. Choi, T. Kim, W. Yu
Performance Evaluation Of Ransac Family
BMVC 2009
-  De la Torre, F. and M. Black,
Robust principal component analysis for computer vision,
Int. Conf. on Computer Vision, ICCV-2001, Vancouver.
-  M. A. Fischler and R. C. Bolles.
Random Sample Consensus: A Paradigm for Model Fitting with
Applications to Image Analysis and Automated Cartography.
Communications of the ACM, 24 (6), 1981.



P. Rousseeuw and A. Leroy.

Robust Regression and Outlier Detection.

Wiley Series in Probability and Mathematical Statistics. Wiley, 1987.



H. Wang and D. Suter.

Robust adaptive-scale parametric model estimation for computer vision

IEEE Transactions on Pattern Analysis and Machine Intelligence, (26),1459–1474, 2004.



Z. Zhang.

Parameter Estimation Techniques; A Tutorial with Application to Conic Fitting.

Rapport inria, 2676, centre de sophia antipolis, 1995.