Thème 1

Moindres carrés : OLS-ODR-TLS

Département Sciences du Numérique
Janvier 2018

1.1 La notation «boîte noire»

On considère un système [1] qui se prête à des mesures sur deux types de grandeurs à valeurs dans $\mathbb{R}$. D'une part, les **covariables** ou **régresseurs** ou variables **explicatives** ou **indépendantes** ou **en entrée** peuvent être fixées et le système réagit à leur valeur. D'autre part, les **réponses** ou variables **expliquées** ou **dépendantes** ou **en sortie** sont telles que le système produit leur valeur, on peut les mesurer. Au système est associé un modèle mathématique paramétré (la fonction f). On souhaite déterminer la valeur des **paramètres** à partir de mesures des variables –**les observations**– réalisées au cours de plusieurs (n) expériences.

Formellement, nous notons:

- les variables en entrée $x \in \mathbb{R}^m$;
- les variables en sortie $y \in \mathbb{R}^q$;
- les paramètres $\beta \in \mathbb{R}^p$;
- le modèle

$$f: \begin{array}{ccc} \mathbb{R}^m \times \mathbb{R}^p & \longrightarrow & \mathbb{R}^q \\ (x, \beta) & \longmapsto & y \end{array}$$

[1] Le système peut être physique, biologique, financier, complexe voire même chimique.

En analyse de données, on parle aussi de modèle de **régression** au sens où la fonction f et ses paramètres permettent une réduction des données d'un phénomène complexe en vue de les représenter par une loi simplificatrice. En régression, nous rassemblons toutes les informations utiles à la modélisation d'un système sous la forme synthétique graphique de la **boîte noire**:

$$x \in \mathbb{R}^m \rightarrow \boxed{\begin{array}{c} \beta \in \mathbb{R}^p \\ \hline m, n, p, q \end{array}} \rightarrow y = f(x, \beta) \in \mathbb{R}^q$$

À cette boîte noire, nous adjoignons le résultat des mesures issues de chaque expérience. Pour l'expérience i ($i \in [\![1, n]\!]$) ce sont la valeur $\widetilde{x}_i$ de la variable x et la valeur $\widetilde{y}_i$ de la variable y. Ce procédé de calcul des β, faisant intervenir n observations des variables x et y est connu sous le nom d'**estimation de paramètres**.

1.2 Modèles de régression et d'erreurs

Dire que le modèle est linéaire est un abus de langage pour désigner le cas où f est affine en β. C'est un cas simple et de référence auquel on se ramène souvent. Le modèle peut s'écrire sous la forme

$$f(x, \beta) = C(x)\beta + D(x)$$

avec

$$C : \mathbb{R}^m \longrightarrow \mathbb{R}^{q \times p}$$
$$D : \mathbb{R}^m \longrightarrow \mathbb{R}^q$$

Dans la suite, on supposera pour simplifier le message que $m = q = 1$ (cas unidimensionnel). La mesure des variables est sujette à des erreurs dont le processus d'estimation doit tenir compte. Ces erreurs peuvent apparaître sur les entrées et/ou les sorties. Dans un premier temps on considère des erreurs sur les sorties uniquement. On perturbe la valeur exacte y_i selon $\widetilde{y}_i = y_i + \varepsilon_i$.

Les erreurs (additives) sont souvent supposées gaussiennes. Dans tous les cas, deux hypothèses d'indépendance (au sens des variables aléatoires qui les modélisent) sous-tendent ces modèles d'erreur: indépendance d'une expérience à l'autre et indépendance entre toute erreur en entrée et toute erreur en sortie. Les erreurs gaussiennes apparaissent souvent quand des facteurs multiples incontrôlés s'accumulent pour perturber faiblement la mesure.

1.3 Moindres carrés ordinaires

Ces méthodes cherchent à rendre « aussi vraie que possible » l'approximation sous-jacente au modèle de régression. Pour cela, elles calculent β qui minimise « conjointement » tous les

$$\|\widetilde{y}_i - f(\widetilde{x}_i, \beta)\|^2 \qquad \forall i \in [\![1, n]\!]$$

Cette expression basique des **résidus** sera plus tard modifiée pour intégrer une pondération des mesures et prendre en compte les erreurs sur les entrées.

Moindres carrés ordinaires (*Ordinary Least Squares*, OLS)

On parle aussi de « régression en distance verticale ». L'estimateur est:

$$\widehat{\beta_{\mathrm{OLS}}} = \underset{\beta \in \mathbb{R}^p}{\operatorname{argmin}} \sum_{i=1}^{n} \|\widetilde{y_i} - f(\widetilde{x_i}, \beta)\|^2 \tag{1.1}$$

1.3.1 Résolutions pour le cas unidimensionnel

Moindres carrés ordinaires

Si f est linéaire en β, le problème OLS se ramène à:

$$\widehat{\beta_{\mathrm{OLS}}} = \underset{\beta \in \mathbb{R}^p}{\operatorname{argmin}} \|A\beta - b\|^2 \tag{1.2}$$

où

$$A = \begin{bmatrix} C(\widetilde{x_1}) \\ \vdots \\ C(\widetilde{x_n}) \end{bmatrix} \in \mathbb{R}^{n \times p} \text{ et } b = \begin{bmatrix} \widetilde{y_1} - D(\widetilde{x_1}) \\ \vdots \\ \widetilde{y_n} - D(\widetilde{x_n}) \end{bmatrix} \in \mathbb{R}^n$$

Une méthode de résolution des équations normales, vue en cours, utilise la pseudo inverse A^+ de A (que l'on peut exprimer via la SVD).

$$\widehat{\beta_{\mathrm{OLS}}} = A^+ b$$

Moindres carrés totaux

Une formulation aux moindres carrés totaux (*Total Least Squares*) est également possible grâce à la SVD. Il s'agit de résoudre :

$$\min_{\beta \in \mathbb{R}^p} \|A\beta - b\|^2$$

On peut le réécrire :

$$A\beta - b = [A \;\; b] \begin{bmatrix} \beta \\ -1 \end{bmatrix} = 0$$

On cherche donc un élément du noyau de $[A \;\; b]$ de la forme $\begin{bmatrix} \beta \\ -1 \end{bmatrix}$. Si le rang de $[A \;\; b]$ est $p+1$, on passe au rang p :

$$[\hat{A} \;\; \hat{b}] = [A \;\; b] - \sigma_{p+1} u_{p+1} v_{p+1}^T$$

$\widehat{\beta_{TLS}}$ est associé à v_{p+1} mais doit être de la forme $\begin{bmatrix} \beta \\ -1 \end{bmatrix}$. Soit $v_{p+1} = \begin{bmatrix} v_{1\ p+1} \\ \vdots \\ v_{p+1\ p+1} \end{bmatrix} \in \mathbb{R}^{p+1}$ d'où

$$\widehat{\beta_{TLS}} = \frac{1}{-v_{p+1\ p+1}} \begin{bmatrix} v_{1\ p+1} \\ \vdots \\ v_{p\ p+1} \end{bmatrix}.$$

1.4 Exercice : transformations géométriques 2D

Les opérations géométriques ont pour but de modifier la position des informations contenues dans l'image sans modifier le niveau de gris. Ces opérations peuvent s'appliquer à la totalité des points d'une image (les pixels), un objet particulier de l'image, voire à certains points spécifiques (recalage de points caractéristiques).

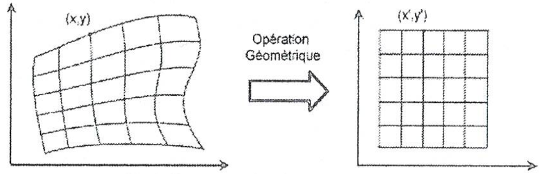
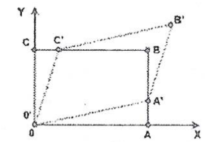


Fig 1 - Recalage géométrique d'une image

Les opérations de base permettent de réaliser des transformations géométriques simples. Elles sont visibles lorsqu'elles affectent une structure spécifique (élément carré, maillage). On considère la transformation suivante :

Cisaillement double /transvection		$\begin{cases} x' = x + \alpha y \\ y' = \beta x + y \end{cases}$

Questions

On souhaite modéliser le problème consistant à trouver la transvection liant 2 ensembles de points dans $\mathbb{R}^2$. On définit $\gamma \in \mathbb{R}^2$ le vecteur de paramètres inconnus défini par : $\gamma = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$.

$$U = \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2 \longrightarrow \boxed{\gamma \in \mathbb{R}^2} \longrightarrow U' = \begin{bmatrix} x' \\ y' \end{bmatrix} = f(U, \gamma) \in \mathbb{R}^2$$

1. Déterminer la matrice $C \in \mathbb{R}^{2\times 2}$ et le vecteur $d \in \mathbb{R}^2$ de la fonction $f(U, \gamma)$ définie sous la forme :
$$f(U, \gamma) = C\,\gamma + d$$
2. Ecrire le problème d'optimisation pour définir le vecteur de paramètres γ en considérant un ensemble de n points transformés par la même transvection.
3. Calculer la solution théorique de ce problème d'optimisation.

1.5 ODR et Moindres carrés totaux

On considère désormais des erreurs à la fois sur les données d'entrée et sur des données sorties. On note $i \in [\![1, n]\!]$ l'indice de l'expérience. Les erreurs sur les entrées perturbent la valeur exacte x_i $\widetilde{x}_i = x_i + \delta_i$ sont complétées par des erreurs sur les sorties: $\widetilde{y}_i = y_i + \varepsilon_i$.

Moindres carrés orthogonaux (*Orthogonal Distance Regression*, ODR)

Les entrées et les sorties sont entachées d'erreurs. Il s'agit d'estimer « conjointement » β et les erreurs en entrée δ_i:

$$P_{\text{ODR}} \left\{ \begin{array}{l} \min \sum_{i=1}^{n} \|\widetilde{y}_i - f(\widetilde{x}_i - \delta_i, \beta)\|^2 + \|\delta_i\|^2 \\ \beta \in \mathbb{R}^p, \delta_i \in \mathbb{R}^m, i \in [\![1, n]\!] \end{array} \right. \tag{1.3}$$

On parle de « régression en distance orthogonale » (*Orthogonal Distance Regression*, ODR).

Moindres carrés totaux

Le problème ODR admet une solution analytique dans le cadre des moindres carrés totaux (*Total Least Squares*), TLS. C'est un cas restreint où les dimensions vérifient $m = p$ et $q = 1$ et le modèle f est bilinéaire de la forme $f(x, \beta) = x^\top \beta + q_0^\top \beta + q_1$ avec $q_0 \in \mathbb{R}^p$ et $q_1 \in \mathbb{R}$. Si $f(\widetilde{x}_i - \delta_i, \bar{\beta}) = \widetilde{y}_i - \varepsilon_i, \forall i \in [\![1, n]\!]$ alors en exploitant les linéarités et en synthétisant les égalités sous forme matricielle:

$$\left(\underbrace{\begin{bmatrix} \widetilde{x}_1{}^\top + q_0^\top \\ \vdots \\ \widetilde{x}_n{}^\top + q_0^\top \end{bmatrix}}_{A} + \underbrace{\begin{bmatrix} -\delta_1^\top \\ \vdots \\ -\delta_n^\top \end{bmatrix}}_{E} \right) \bar{\beta} = \underbrace{\begin{bmatrix} \widetilde{y}_1 - q_1 \\ \vdots \\ \widetilde{y}_n - q_1 \end{bmatrix}}_{b} + \underbrace{\begin{bmatrix} -\varepsilon_1 \\ \vdots \\ -\varepsilon_n \end{bmatrix}}_{r}$$

Par suite, l'équation (1.3) s'écrit:

$$P_{\text{TLS}} \left\{ \begin{array}{l} \min \|[E \quad r]\|^2 \\ \beta \in \mathbb{R}^p, E \in \mathbb{R}^{n \times p}, r \in \mathbb{R}^n \\ (A + E)\beta = b + r \end{array} \right.$$

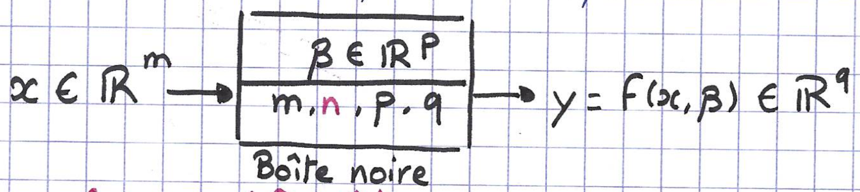
La résolution de (P_{TLS}) se base sur la décomposition en valeurs singulières (SVD) de $[A \quad b]$.

CSAD
CTD 3.1

Analyse de données : Moindres carrés

Modélisation paramétrique

- variables en entrée : $x \in \mathbb{R}^m$ ← données
- variables en sortie : $y \in \mathbb{R}^q$
- des paramètres $\beta \in \mathbb{R}^p$ ← données
- Un modèle $f : \begin{cases} \mathbb{R}^m \times \mathbb{R}^p \longrightarrow \mathbb{R}^q \\ (x, \beta) \longrightarrow f(x,\beta) \end{cases}$ ← données



- n expériences $(\hat{x}_i, \hat{y}_i)$

1/ f linéaire en β (ou après linéarisation)

$$f(x,\beta) = C(x)\beta + D(x)$$

avec $C : \mathbb{R}^n \longrightarrow \mathbb{R}^{q\times p}$
$D : \mathbb{R}^m \longrightarrow \mathbb{R}^q$

exemple : transformation de points de $\mathbb{R}^2$

$$x_i = (u_i, v_i) \longrightarrow (u_i', v_i') = y_i'$$

$$\begin{bmatrix} u_i' \\ v_i' \end{bmatrix} = \begin{bmatrix} a_1 & a_2 \\ b_1 & b_2 \end{bmatrix} \begin{bmatrix} u_i \\ v_i \end{bmatrix} + \begin{bmatrix} a_3 \\ b_3 \end{bmatrix}$$

○ paramètres

$$f(x,\beta) = C(x)\beta + D(x)$$

Posons $\beta = [a_1, a_2, a_3, b_1, b_2, b_3]^T \in \mathbb{R}^6$ $\quad C \in \mathbb{R}^{2\times 6}$

$$\begin{bmatrix} u_i' \\ v_i' \end{bmatrix} = \begin{bmatrix} u_i & v_i & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & u_i & v_i & 1 \end{bmatrix} \beta + 0_{\mathbb{R}^2}$$

$$A = \begin{bmatrix} u_1 & v_1 & 1 & & & \\ u_2 & v_2 & 1 & & 0 & \\ & \vdots & & & & \\ & & & u_1 & v_1 & 1 \\ & 0 & & u_2 & v_2 & 1 \\ & & & & \vdots & \end{bmatrix} \quad \text{et} \quad b = \begin{pmatrix} u_1' \\ u_2' \\ \vdots \\ v_1' \\ v_2' \\ \vdots \end{pmatrix}$$

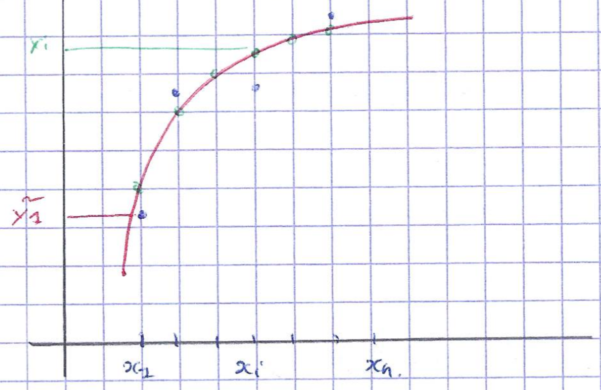
$$A\beta = b$$

- Cas idéal $A\beta = b$
- Réalité : les données sont des données mesurées avec de l'imprécision (erreu

⇒ 3 types de situations :
- erreur sur les données en sortie
- erreur sur les données en entrée et en sortie
- erreur sur les données en entrée.

Problèmes aux moindres carrés (OLS)

Hypothèses : $\hat{x}_i = x_i$ $\mathbb{R}$
$\hat{y}_i = y_i + \Delta y_i$ $\mathbb{R}$.



$y_i = f(x_i, \beta)$ $\forall i \in \{1, n\}$.

β minimise l'erreur (norme du résidu)

$e_i = \hat{y}_i - y_i = \hat{y}_i - f(x_i, \beta)$

$\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{n} \| \hat{y}_i - f(x_i, \beta) \|^2$

$\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{n} \| \underbrace{\Delta y_i}_{\text{dstce verticale}} \|^2$

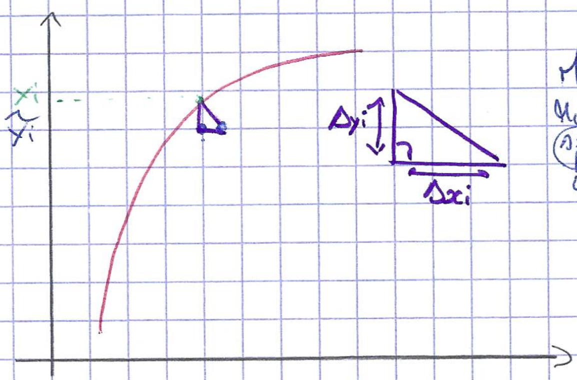
Pb de regression en distance verticale
On va chercher la courbe (fct β) qui minimise la norme des résidus

$\| \hat{y}_i - f(x_i, \beta) \| = \| \Delta y_i \|$ (dstce entre $(x_i, \hat{y}_i)$ et la courbe).

$\Rightarrow$ Solution analytique.

Moindres carrés orthogonaux

Hypothèses : $\hat{x}_i = x_i + \Delta x_i$ $\mathbb{R}$
$\hat{y}_i = y_i + \Delta y_i$ $\mathbb{R}$.



Min d_i^2
$\min_{\Delta x_i, \beta} \Delta x_i^2 + \Delta y_i^2$.
↳ pas de solution analytique
(Schéma itératif pour trouver une sol.).

Estimer simultanément $\hat{\beta}$ et $\widehat{\Delta x_i}$ tq.

$$\min_{\beta, \Delta x_i} \sum_{i=1}^{n} \| f(\hat{x}_i - \Delta x_i, \beta) - \hat{y}_i \|^2 + \| \Delta x_i \|^2$$

minimisation de la distance orthogonale

CSAD
CTD 3.2

Moindres carrés sur les données (en entrée)

Hypothèses : $\hat{x_i} = x_i + \Delta x_i$
$\hat{y_i} = y_i$.

$$\min_{\beta, \Delta x_i} \sum_{i=1}^{n} \|\Delta x_i\|^2$$

Résolution des moindres carrés ordinaires (unidimensionnel)

$$\min_{\beta} \sum_{i=1}^{n} \|\hat{y_i} - f(x_i, \beta)\|^2 \qquad (P)$$

f linéaire en β.

$$\min_{\beta} \frac{1}{2} \|A\beta - b\|^2 \qquad (P')$$ avec $R(\beta) = \frac{1}{2}\|A\beta - b\|^2$.

$$\boxed{CN : \nabla R(\beta) = 0}$$

$R(\beta) = \frac{1}{2}(A\beta - b \mid A\beta - b)$

$\nabla R(\beta) = (A \mid A\beta - b) = A^T(A\beta - b)$.

$\nabla R(\beta) = 0$ $\quad A^T A\hat{\beta} - A^T b = 0$
$A^T A \hat{\beta} = A^T b$.

$A^T A$ est inversible donc

$$\hat{\beta} = (A^T A)^{-1} A^T b.$$
$$= \underbrace{A^+}_{\text{pseudo-inv.}} b.$$

Exercice : transformations géométriques 2D.

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} y & 0 \\ 0 & x \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + \begin{bmatrix} x \\ y \end{bmatrix} = f(v, \gamma)$$

$C = \begin{bmatrix} y & 0 \\ 0 & x \end{bmatrix}$ $\quad d = \begin{pmatrix} x \\ y \end{pmatrix}$

$$R(\gamma) = U' - f(U,\gamma).$$

$$U = \begin{pmatrix} x_1 \\ \vdots \\ x_n \\ y_1 \\ \vdots \\ y_n \end{pmatrix} \quad R(\gamma) = b - A\gamma$$

$$A = \begin{pmatrix} y_1 & 0 \\ \vdots & \vdots \\ y_n & 0 \\ 0 & x_1 \\ \vdots & \vdots \\ 0 & x_n \end{pmatrix} \quad b = \begin{pmatrix} x_1' - x_1 \\ \vdots \\ x_n' - x_n \\ y_1' - y_1 \\ \vdots \\ y_n' - y_n \end{pmatrix}.$$

$$\hat{\gamma} = (A^T A)^{-1} A^T b. \qquad \underset{2\times 2}{A^T A} = \begin{bmatrix} \sum y_i^2 & 0 \\ 0 & \sum x_i^2 \end{bmatrix}.$$

tous les x_i et y_i ne st pas nulles,

Donc $A^T A$ inversible.

et $(A^T A)^{-1} = \begin{pmatrix} \frac{1}{\sum y_i^2} & 0 \\ 0 & \frac{1}{\sum x_i^2} \end{pmatrix}.$

$$A^T b = \begin{bmatrix} \sum_{i=1}^{n} y_i (x_i' - x_i) \\ \sum_{i=1}^{n} x_i (y_i' - y_i) \end{bmatrix}.$$

$$(A^T A^{-1})(A^T b) = \begin{pmatrix} \frac{\sum y_i (x_i' - x_i)}{\sum y_i^2} \\ \frac{\sum x_i (y_i' - y_i)}{\sum x_i^2} \end{pmatrix} = \hat{\gamma}.$$

Resolution d'un problème aux moindres carrés totaux.

formulation implicite (je fais "disparaitre" les sorties).

$$\min_{\|\bar\beta\|=1} \left(\frac{1}{2}\right) \|\bar A \bar\beta\|^2$$

dans l'idéal ce mini est atteint pour $\bar A \bar\beta = 0$.

$\bar\beta \neq 0$ tq $\bar\beta \in \ker \bar A$

<u>SVD</u> $\bar A = U \Sigma V^T$

$\bar A \in \mathbb{R}^{q\times p}$ $\quad \bar A = \sum_{i=1}^{p} \sigma_i u_i v_i^T \xrightarrow{\text{rang } p-1} \hat{\bar A} = \sum_{i=1}^{p-1} \sigma_i u_i v_i^T.$

rang p. $\qquad \bar\beta = v_p.$

$\min_\beta \|A\beta - b\|^2 \qquad A\beta - b = [A\ b]\begin{bmatrix} \beta \\ -1 \end{bmatrix}$

SVD $\bar A$ $\qquad \bar\beta = v_{p+1}$

$\bar\beta = \begin{bmatrix} \beta \\ -1 \end{bmatrix} \quad \begin{bmatrix} -v_{1\,p+1} \end{bmatrix} \qquad \beta = \left[\frac{v_{1\,p+1}}{v_{p+1\,p+1}} \right].$

CSAD
CTD 3.3

$f'(\lambda) = \lambda (r^{(p)})^T A r^{(p)} - (r^{(p)})^T r^{(p)}$

λ_p tq $f'(\lambda_p) = 0$ $\quad \lambda_p = \dfrac{(r^{(p)})^T r^{(p)}}{(r^{(p)})^T A r^{(p)}}$

① $r^{(p+1)} \perp r^{(p)}$

$(r^{(p+1)} \mid r^{(p)}) = \cancel{(r^{(p)}|r^{(p)})} - \lambda_p (Ar^{(p)} \mid r^{(p)}) = 0$ étant donné que

$\lambda_p =$

(λ, μ) couple propre de A

$x^{(0)} = x^* - \beta u$

$x^{(1)} = x^*$?

$x^{(1)} = x^{(0)} + \lambda_0 r^{(0)}$

$x^{(1)} = x^{(0)} + \dfrac{\| b - Ax^{(0)} \|^2}{(b - Ax^{(0)})^T A (b - Ax^{(0)})} (b - Ax^{(0)})$

$x^{(1)} = x^* - \beta u + \dfrac{\| b - \cancel{Ax^*} + A\beta u \|^2}{(0 - \beta \lambda u)^T A (-\beta \lambda u)} (-\beta \lambda u)$

$= x^* - \beta u + \dfrac{\lambda^2 \beta^2 \|u\|^2}{\beta^2 \lambda^3 \|u\|^2} (-\beta \lambda u)$

$= x^* - \beta u + \beta u$

$x^{(1)} = x^*$