Détection d'Erreurs de Mesure par Equilibrage

par
Paul Deheuvels
paul.deheuvels@upmc.fr

1 Introduction.

Nous modélisons un système industriel par un graphe, composé d'un ensemble $\{j = 1, \ldots, n\}$ de noeuds, et d'un ensemble $\{i = 1, \ldots, m\}$ d'arcs. Chaque arc transfert une quantité variable de matière d'un noeud à un autre. La masse transitant par un arc est mesurée par un débitmètre, dont les cumuls évaluent le flux transféré dans une période de référence. Un noeud est un point de concentration, auquel parviennent les flux des arcs entrants (notés positivement), et duquel sont issus les flux des arcs sortants (notés négativement). Chaque noeud doit avoir un bilan matière neutre, le total des masses entrantes (dans une période de référence) devant être égal à celui des masses sortantes (dans la même période). Ceci se traduit par une équation, pour chaque noeud:

$$\sum(\text{masses entrantes}) - \sum(\text{masses sortantes}) = 0. \tag{1.1}$$

La mesure X (le flux observé) de la masse μ (le flux exact) des matières ayant transité par un arc dans une période donnée n'est pas exacte, le débitmètre ne fournissant, après cumul, qu'une mesure approchée X de μ, entachée d'une erreur aléatoire. Nous supposerons ici que le flux observé X suit une loi de probabilité normale $N(\mu, \sigma^2)$, de moyenne égale au flux exact μ, et de variance σ^2 connue. Cette relation est notée $X \stackrel{d}{=} N(\mu, \sigma^2)$. Ici, X est une quantité observée, alors que μ est une quantité inconnue. On supposera que σ^2, paramètre caractérisant la précision des cumuls d'observations fournies par le débitmètre, est connu. Cette hypothèse pourra être modifiée dans des modèles plus complexes, de même que la modélisation de X par une loi normale.

Au cours d'une période de fonctionnement, les flux cumulés $\{X_i : i = 1, \ldots, n\}$ des arcs composent un ensemble de variables aléatoires normales, supposées mutuellement indépendantes, et telles que $X_i \stackrel{d}{=} N(\mu_i, \sigma_i^2)$ pour chaque $i = 1, \ldots, n$. On range ces observations dans un vecteur de l'espace $\mathbb{R}^n$, noté

$$\mathbf{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix} \stackrel{d}{=} N_n(\boldsymbol{\mu}, \Sigma), \quad \text{où} \quad \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix} \in \mathbb{R}^n, \quad \text{et où} \quad \Sigma = \begin{bmatrix} \sigma_1^2 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \sigma_n^2 \end{bmatrix}, \tag{1.2}$$

est une matrice de variances-covariances $n \times n$. L'hypothèse d'indépendance entre les mesures impose que cette matrice soit diagonale, mais, bien entendu, d'autres configurations sont possibles. Ici, $N_n(\boldsymbol{\mu}, \Sigma)$ désigne une loi normale dans $\mathbb{R}^n$, de moyenne $\boldsymbol{\mu}$ et de matrice de variances-covariances Σ. L'équation (1.1), traduisant la loi de conservation de la matière pour le noeud $j = 1, \ldots, m$ s'écrit alors sous la forme

$$\sum_{i=1}^{n} a_{j,i}\mu_i = 0, \tag{1.3}$$

où

$$a_{j,i} = \begin{cases} 1 & \text{si l'arc } i \text{ apporte une quantité de matière positive au noeud } j; \\ -1 & \text{si l'arc } i \text{ extrait une quantité de matière positive du noeud } j; \\ 0 & \text{si l'arc } i \text{ n'est pas relié au noeud } j. \end{cases} \tag{1.4}$$

Avec ces notations et hypothèses, le système se résume aux formules matricielles

$$\mathbf{X} \stackrel{d}{=} N_n(\boldsymbol{\mu}, \Sigma), \quad \text{avec} \quad A\boldsymbol{\mu} = \mathbb{O}_m := \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \in \mathbb{R}^m, \quad \text{et où} \quad A := \begin{bmatrix} a_{1,1} & \ldots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{m,1} & \ldots & a_{m,n} \end{bmatrix}. \tag{1.5}$$

Le vecteur de mesures $\mathbf{X} \in \mathbb{R}^n$ ne fournit pas un *bilan équilibré*, au sens qu'il ne satisfait pas les relations de conservation des masses, du fait que $A\mathbf{X} \neq \mathbb{O}_m$. On corrige les observations $\mathbf{X}$, en tenant compte de la relation admise $A\boldsymbol{\mu} = \mathbb{O}_m$, pour obtenir un vecteur corrigé $\widehat{\boldsymbol{\mu}}$, qui constitue l'estimation par la méthode du maximum de vraisemblance de $\boldsymbol{\mu}$, et vérifie la relation d'équilibre $A\widehat{\boldsymbol{\mu}} = \mathbb{O}_m$. Ce vecteur est donnée par

$$\widehat{\boldsymbol{\mu}} = \mathbf{X} - \Sigma A' \left(A\Sigma A'\right)^{-1} A\mathbf{X}. \tag{1.6}$$

Cette relation est établie, sous la réserve que les matrices $A\Sigma A'$ et Σ, de dimension $n \times n$, soient inversibles. Si ce n'est pas le cas, on peut faire usage de pseudo-inverses pour obtenir une relation valide sans cette restriction. Dans la formule ci-dessus, la notation M' désigne la transposée de la matrice M, de même que la notation M^{-1} désigne l'inverse de M lorsque celui-ci a un sens. L'estimation $\widehat{\boldsymbol{\mu}}$ de $\boldsymbol{\mu}$ est centrée, au sens que

$$\mathbb{E}(\widehat{\boldsymbol{\mu}}) = \boldsymbol{\mu} - \Sigma A' \left(A\Sigma A'\right)^{-1} A\boldsymbol{\mu} = \boldsymbol{\mu}. \tag{1.7}$$

Les vecteurs aléatoires $\widehat{\boldsymbol{\mu}}$ et $\mathbf{X} - \widehat{\boldsymbol{\mu}}$ sont, à la fois orthogonaux (en tant que vecteurs), relativement au produit scalaire $\langle u, v\rangle_{\Sigma^{-1}} = u'\Sigma^{-1}v$, et mutuellement indépendants. On vérifie, en effet, d'une part, que

$$\begin{aligned} \left(\mathbf{X} - \widehat{\boldsymbol{\mu}}\right)' \Sigma^{-1} \left(\widehat{\boldsymbol{\mu}}\right) &= \mathbf{X}'A' \left(A\Sigma A'\right)^{-1} A\Sigma \times \Sigma^{-1} \times \left\{\mathbf{X} - \Sigma A' \left(A\Sigma A'\right)^{-1} A\mathbf{X}\right\} \\ &= \mathbf{X}'A' \left(A\Sigma A'\right)^{-1} A\mathbf{X} - \mathbf{X}'A' \left(A\Sigma A'\right)^{-1} A\mathbf{X} = 0, \end{aligned} \tag{1.8}$$

et, d'autre part, que

$$\begin{aligned} &\mathbb{E}\Big(\Big\{\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}\Big\}\Big\{\mathbf{X} - \widehat{\boldsymbol{\mu}}\Big\}'\Big) \\ &= \mathbb{E}\left(\left\{\Big\{\mathbf{X} - \boldsymbol{\mu}\Big\} - \Sigma A' \left(A\Sigma A'\right)^{-1} A\Big\{\mathbf{X} - \boldsymbol{\mu}\Big\}\right\} \times \Big\{\mathbf{X} - \boldsymbol{\mu}\Big\}' A' \left(A\Sigma A'\right)^{-1} A\Sigma\right) \\ &= \Sigma A' \left(A\Sigma A'\right)^{-1} A\Sigma - \Sigma A' \left(A\Sigma A'\right)^{-1} A\Sigma = \mathbb{O}_{n,n}. \end{aligned} \tag{1.9}$$

Ici, $\mathbb{E}(\cdot)$ désigne l'espérance, et $\mathbb{O}_{n,n}$ la matrice $n \times n$ composée de 0. Par ailleurs, la matrice de variances-covariances de $\mathbf{X} - \widehat{\boldsymbol{\mu}}$ est donnée par

$$\mathbb{E}\left(\Big\{\mathbf{X} - \widehat{\boldsymbol{\mu}}\Big\}\Big\{\mathbf{X} - \widehat{\boldsymbol{\mu}}\Big\}'\right) = \Sigma A' \left(A\Sigma A'\right)^{-1} A\Sigma =: \begin{bmatrix} \gamma_{1,1} & \cdots & \gamma_{1,n} \\ \vdots & \ddots & \vdots \\ \gamma_{n,1} & \cdots & \gamma_{n,n} \end{bmatrix}. \tag{1.10}$$

2 Détection d'un débitmètre faussé.

Avec les notations du §1, la correction d'équilibrage du i$^{\text{ème}}$ débitmètre est donnée par $\delta_i = \widehat{\mu}_i - X_i$, où

$$\widehat{\boldsymbol{\mu}} = \begin{bmatrix} \widehat{\mu}_1 \\ \vdots \\ \widehat{\mu}_n \end{bmatrix} \stackrel{d}{=} N_n\left(\boldsymbol{\mu}, \Sigma A' \left(A\Sigma A'\right)^{-1} A\Sigma\right). \tag{2.11}$$

Compte tenu de (1.10), pour le i$^{\text{ème}}$ débitmètre, la correction d'équilibrage suit une loi normale de paramètres

$$\delta_i \stackrel{d}{=} N(0, \gamma_{i,i}). \tag{2.12}$$

On décide alors que la i^{ème} mesure est fausse si $|\delta_i|/\sqrt{\gamma_{i,i}}$ dépasse le seuil critique à 97.5% d'une loi normale standard (soit, approximativement, 1.96).

La procédure ci-dessus est idéale si on connaît à l'avance le débitmètre à contrôler. Si on l'applique pour chacun des n débitmètres, il faudra adapter le seuil de détection (ici, fixé à 5%) pour tenir compte de la répétition des tests de détection effectués. Il s'agit ici de l'hypothèse la plus simple, où un seul débitmètre peut tomber en panne. La méthodologie pour détecter plusieurs pannes simultanées de débitmètres nécessite des modifications de cette procédure, à partir de ce modèle de base.

- - - - - - - - - - - - - - -