NONLINEAR MOBILE SENSOR CALIBRATION USING INFORMED SEMI-NONNEGATIVE MATRIX FACTORIZATION WITH A VANDERMONDE FACTOR

Clément Dorffer, Matthieu Puigt, Gilles Delmaire, and Gilles Roussel

Univ. Littoral Côte d’Opale, EA 4491 – LISIC, F-62228 Calais, France

ABSTRACT
In this paper we aim to blindly calibrate a mobile sensor network whose sensor outputs and the sensed phenomenon are linked by a polynomial relationship. The proposed approach is based on a novel informed semi-nonnegative matrix factorization with a Vandermonde factor matrix. The proposed approach outperforms a matrix-completion-based method in a crowdsensing-like simulation of particulate matter sensing.

Index Terms—nonlinear sensor calibration, informed semi-NMF, structured matrix factorization

1. INTRODUCTION
Crowdsensing consists of acquiring some geolocalized and time-stamped data from a crowd of mobile sensors provided by mobile devices, e.g., smartphones [1]. However, crowdsensing sensors are usually low cost and must be oftenly calibrated in the wild, as it may not be possible to request them to regularly go to a laboratory. To deal with such an issue, specific blind sensor calibration methods have been proposed. Such techniques are usually divided into two main families, namely micro- and macro-calibration. Micro-calibration consists of independently calibrating each sensor of a network while macro-calibration operates on the whole set of sensors [2]. Most blind calibration approaches allow to tackle gain [3, 4], gain/offset [5, 6, 7, 8, 9], gain/phase [10, 11] calibration, and sensor drift [12]. When the sensors of the network can move, Blind Mobile Sensor Calibration (BMSC) techniques are usually exploiting the rendezvous model [13] which assumes that sensors in the same spatio-temporal vicinity should acquire the same data. Contrary to [12], the authors in [6, 7, 8, 9] consider the situation when some sensors are perfectly calibrated, thus helping the calibration of the rest of the sensor network. In particular, the work in [6, 7] proposes a multi-hop micro-calibration structure. That is, calibrated sensors are used to calibrate uncalibrated ones which lie in the same vicinity. These sensors are then considered as calibrated and used to calibrate other ones when they move. The operation is then repeated until any sensor in the network is calibrated. On the contrary, our previous work in [8, 9] is based on macro-calibration and is revisiting BMSC as an informed Nonnegative Matrix Factorization (NMF) problem. Our approaches are assuming a nonnegative affine calibration function, which is valid for a limited number of sensors only.

In this paper, we extend our previous work by proposing a novel nonlinear calibration method where the nonlinear function is assumed to be polynomial. Interestingly, while being extremely general, such a model for BMSC has been considered in a very limited number of papers, i.e., [14, 15]. The work in [14] also considers a polynomial calibration function and proposes two BMSC approaches based on nullspace—i.e., an extension of [5]—and on moments, respectively. The work in [15] considers an extended moment-based calibration method, where the nonlinear calibration function is assumed to be piecewise linear. However, it should be noticed that the nullspace-based method is very sensitive to noise [14] and needs some knowledge on the true subspace where the sensed data should lie. The moment-based calibration approach is more robust to noise but it needs a very long integration time to perform an accurate calibration [14, 15]. On the contrary, the method we propose in this paper—which is extending our previous work—is based on an informed semi-NMF framework and neither requests the exact knowledge of the sensed data subspace nor needs a long integration time.

The remainder of this paper is organized as follows. We introduce the nonlinear BMSC problem in Section 2 and our proposed approach in Section 3. Its experimental performance is investigated in Section 4 while we conclude and discuss about future work in Section 5.

2. PROBLEM STATEMENT
In this paper we consider a set of $m$ heterogeneous, geolocalized, time-stamped, mobile, and possibly uncalibrated sensors to provide measurements over a fixed area and over time. We now introduce the definitions used in this paper.

Definition 1 ([13]) A rendezvous is a temporal and spatial vicinity between two sensors.

A rendezvous is characterized by a time interval $\Delta t$, and a spatial distance $\Delta d$ which depend on the observed phenomenon.

Definition 2 ([8]) A scene $S$ is a discretized area observed during a time interval $[t, t+\Delta t)$. The size of the spatial pixels...
are set so that any couple of points inside the same pixel have a distance below $\Delta_0$.

As shown in Fig. 1, a scene is a grid of locations where the sensors go to. When two sensors share the same location in $\mathcal{S}$, they are in rendezvous and should acquire the same data.

The sensor output $x_{i,j}$—provided by Sensor $j$ at Location $i$ of a scene $\mathcal{S}$—is linked to the corresponding physical value—denoted $y_i$—through a nonlinear calibration function $F_j$ accurately inferred by a polynomial (of degree $N$), i.e.,

$$x_{i,j} \approx F_j(y_i) \approx f_{1,j} \cdot y_i + \cdots + f_{N+1,j} \cdot y_i^N,$$

where the $f_{k,j}$ terms are the calibration parameters associated with Sensor $j$. Please note that such a relationship is extending the affine calibration model [8, 9] as the latter is derived from Eq. (1) by assuming that $f_{k,j} = 0$ for $k \geq 3$.

Assuming a scene composed of $n$ locations and fully observed by $m$ nonlinear mobile sensors (with $\min(n,m) \gg N$), we respectively define the $n \times (N + 1)$ and $(N + 1) \times m$ matrices $G$ and $F$ as

$$G \triangleq \begin{bmatrix} 1 & y_1 & \cdots & y_1^N \\ \vdots & \vdots & \ddots & \vdots \\ 1 & y_m & \cdots & y_m^N \end{bmatrix}, \quad F \triangleq \begin{bmatrix} f_{1,1} & \cdots & f_{1,m} \\ \vdots & \ddots & \vdots \\ f_{N+1,1} & \cdots & f_{N+1,m} \end{bmatrix},$$

and we derive the matrix form of Eq. (1), i.e.,

$$X \approx G \cdot F,$$

where $X \triangleq [x_{i,j}]_{i,j}$ is a low-rank well-conditionned matrix. $G$ is a Vandermonde matrix—denoted $G \triangleq \text{VDM}(y,N)$ below—whose columns provide the monomial expansion of the sensed physical values $y \triangleq [y_1, \ldots, y_m]^T$ while $F$ contains the calibration parameters.

It should be noticed that estimating both $G$ and $F$ from $X$ is a specific nonlinear matrix factorization problem, which has never been investigated to the best of the authors’ knowledge. Indeed, most nonlinear matrix factorization approaches were proposed for bilinear source separation [16, 17, 18], whose nonlinear dependencies in one matrix factor are different from the Vandermonde factor considered in this paper.

As the sensors are unconstrained in their moves, none might sense the whole scene, thus leading to missing values which are taken into account by applying a binary mask $\bar{W}$—informing the presence/absence of data in $X$—on the factorization, i.e., denoting $\circ$ the Hadamard product:

$$W \circ X \approx W \circ (G \cdot F),$$

Lastly, we assume that $X$ and $G$ are nonnegative, which is a valid assumption in many environmental applications\(^2\), Moreover, we assume that one sensor—say Sensor $m$—is calibrated and that its calibration parameters read

$$f_{2,m} = 1$$

This implies that an available data point in the last column of $X$—say $x_{i,m}$—is equal to the sensed phenomenon $y_i$ in the second column of $G$. Since $G$ is a Vandermonde matrix, its monomial expansions are known too, i.e., the whole $i$-th row of $G$ is known. As a consequence, we get some knowledge on both $G$ and $F$ that may help to enhance the factorization (and in particular, to suppress the scale ambiguity inherent to matrix factorization and that we must avoid to perform sensor calibration). Using the parameterization proposed in [20], we rewrite $G$ and $F$ with respect to their set and free parts, i.e.,

$$G \triangleq \Omega_G \circ \Phi_G + \Omega_G \circ \Delta_G, \quad F \triangleq \Omega_F \circ \Phi_F + \Omega_F \circ \Delta_F,$$

where $\Omega_G$ (respectively, $\Omega_F$) is a binary matrix informing the presence/absence of constraints in $G$ (respectively, $F$), $\Phi_G$ (respectively, $\Phi_F$) is the matrix containing the set entries in $G$ (respectively, $F$), $\Delta_G$ (respectively, $\Delta_F$) is the matrix of the free parameters in $G$ (respectively, $F$), and $\Omega_G \triangleq \mathbb{1}_{n \times (N+1)} - \Omega_G$ (respectively, $\Omega_F \triangleq \mathbb{1}_{(N+1) \times m} - \Omega_F$), where $\mathbb{1}_{i \times j}$ is the $i \times j$ matrix of ones.

In this paper, we thus aim to estimate $F$ (and $G$) from:

$$\begin{align*}
\{\hat{G}, \hat{F}\} &= \arg\min_{G,F} ||W \circ (X - G \cdot F)||_F^2 \\
\text{s.t. } G &= \text{VDM}(y,N) \geq 0, \\
G &= \Omega_G \circ \Phi_G + \Omega_G \circ \Delta_G, \\
F &= \Omega_F \circ \Phi_F + \Omega_F \circ \Delta_F, 
\end{align*}$$

which is possible, provided enough diversity in $y$ and $F$.

3. PROPOSED APPROACH

As Eq. (7) is non-convex w.r.t. both $G$ and $F$, we split it into

$$\begin{align*}
\hat{G} &= \arg\min_{G} ||W \circ (X - G \cdot F)||_F^2 \\
\text{s.t. } G &= \text{VDM}(y,N) \geq 0, \\
G &= \Omega_G \circ \Phi_G + \Omega_G \circ \Delta_G, 
\end{align*}$$

and

$$\begin{align*}
\hat{F} &= \arg\min_{F} ||W \circ (X - G \cdot F)||_F^2 \\
\text{s.t. } F &= \Omega_F \circ \Phi_F + \Omega_F \circ \Delta_F, 
\end{align*}$$

which then can be alternately and iteratively solved, as explained in Subsections 3.1 and 3.2, respectively.

\(^1\)However [19] considers a Vandermonde factor in tensor factorization.

\(^2\)x_{i,j} and $y_i$ may respectively represent a voltage and a concentration [8].
3.1. Updating strategy for the Vandermonde matrix $G$

From the parameterization (6), it should be noticed that only the free part of $G$, i.e., $\Delta G$, has to be updated in the optimization problem (8). The associated cost function then reads

$$J_G(\Delta G) = \sum_{i=1}^{m} \sum_{j=1}^{n} W_{i,j}^2 \cdot \left( \tilde{X}_{i,j} - ((\Omega_G \circ \Delta G) \cdot F)_{i,j} \right)^2,$$

where $\tilde{X} \triangleq X - (\Omega_G \circ \Phi_G) \cdot F$. As explained above, if a value of $y$ is known, then all the corresponding line of $G$ is set in $\Phi_G$. On the contrary, when a value—say $y_i$—is unknown, then all the $i$-th row of $G$ must be estimated, i.e., the $i$-th row of $\Delta G$ must be updated. Expressing Eq. (10) with respect to $\Delta_y$—the second column of $\Delta G$—reads

$$J_G(\Delta_y) = \sum_{i=1}^{m} \sum_{j=1}^{n} W_{i,j}^2 \cdot \left( \tilde{X}_{i,j} - \sum_{k=0}^{N} (\Omega_G)_{i,k} \cdot (\Delta_y)^k \cdot F_{j,k} \right)^2.$$  

Using this last expression, we propose a two-step update of the whole matrix $G$ which first consists of updating $\Delta_y$, using one iteration of a projected gradient descent—in order to respect the data nonnegativity—and then propagating this updated column into $G$ so as to respect the Vandermonde structure (2). Deriving the cost function (11) yields to

$$\nabla J_G(\Delta_y) = \text{diag} \left[ (W^2 \circ Z) \cdot (F^s)^T \cdot (\tilde{\Omega}_G \circ U \circ \Delta_y^s)^T \right],$$

where $Z \triangleq (\Omega_G \circ \Delta G) \cdot F - \tilde{X}$, $U$ is a $N \times N$ matrix whose $(i,j)$-th element $u_{i,j} \triangleq j$, $W^2 \triangleq W \circ W$, $\tilde{\Omega}_G$ is the matrix composed of the $N-1$ last columns of $\Omega_G$, $\Delta_y^s$ is the matrix composed of the $N-1$ first columns of $\Delta G$ and $F^s$ is the matrix composed of the $N-1$ last rows of $F$. The projected gradient descent update then reads

$$\Delta_y \leftarrow \left[ \Delta_y - \lambda_G \cdot \nabla J_G(\Delta_y) \right]^+, \tag{13}$$

where $\lambda_G$ is the descent step size for $G$ and $[\cdot]^+$ is the projection operator on $\mathbb{R}^+$. 

3.2. Updating strategy for the parameter matrix $F$

Using the parameterization (6), the cost function (9) can be expressed with respect to $\Delta F$ as

$$J_F(\Delta F) = \sum_{i=1}^{m} \sum_{j=1}^{n} W_{i,j}^2 \cdot \left( \tilde{X}_{i,j} - (\Omega_F \circ \Delta F)_{i,j} \right)^2,$$

where $\tilde{X} \triangleq X - (\Omega_F \circ \Phi_F)$. Its gradient reads

$$\nabla J_F(\Delta F) = \tilde{\Omega}_F \circ \left[ G^T \cdot (W^2 \circ (G \circ \tilde{\Omega}_F \circ \Delta F) - \tilde{X}^T) \right]. \tag{15}$$

As $F$ is not subject to any sign constraint, we thus propose to update it by one gradient descent step which leads to

$$\Delta F \leftarrow \Delta F - \lambda_F \cdot \nabla J_F(\Delta F), \tag{16}$$

where $\lambda_F$ is the descent step size for $F$.

3.3. Algorithmic strategies

We here discuss some implementation issues which were met in some preliminary tests. As the proposed method is iterative, it should be initialized. We follow a similar strategy as in [8], i.e., we first apply matrix completion [21] to $X$. Using its completed version, we derive from Eqs. (5) and (2) an estimation of $y$ and $G$, respectively. A naive calibration—which was used to initialize the approach in [8]—consists of estimating $F$ by least squares. However, the semi-NMF with this initialization did not always provide a good enhancement. On the contrary, initializing $F$ with random realizations around known theoretical average values—which are provided by the sensor manufacturer—was found to be a better strategy.

A tricky aspect of the proposed gradient-based method is the choice of the step sizes $\lambda_G$ and $\lambda_F$, respectively introduced in Eqs. (13) and (16). Indeed, if these step sizes are too small, the proposed approach needs a lot of iterations to converge, which is time consuming. On the contrary, if it is too big, the projection in the update of $G$ tends to be massively applied, hence providing a very poor performance along iterations. Moreover, we found in preliminary tests that choosing well-suited step sizes is also depending of the missing data proportion\(^4\). We thus found from preliminary tests a trade-off by setting step sizes varying with the missing value proportion $\rho$ in $X$, i.e.,

$$\lambda_G = 0.001(\rho + 0.1), \quad \lambda_F = 0.01(\rho + 0.1). \tag{17}$$

4. EXPERIMENTAL VALIDATION

In this section, we aim to investigate the enhancement provided by our proposed informed nonlinear Semi-NMF method for BMSC. For that purpose, we simulate a crowdsensing-like particulate matter sensing during a time interval $[t, t + \Delta_t]$, which satisfies the assumptions in Section 2. The scene is a $10 \times 10$ discretized area (the length of $y$ is thus equal to $n = 100$) which is observed by $m = 31$ sensors, i.e., $m-1$ uncalibrated and mobile dust sensors [22] and one calibrated, high quality, and mobile sensor\(^5\).

The observed concentrations in $y$ range from 1e-3 to 1.4 mg/m$^3$, for which the nonlinear sensor response is accurately inferred by a polynomial of degree $N = 2$ [22].

\(^4\)Ideally, the optimal step size should be estimated within iterations. However, such an estimation is non-trivial for the Vandermonde factor and out of the scope of this paper.

\(^5\)Actually, we get $k$ fixed, calibrated, and highly accurate sensors whose outputs are modeled as those of the $m$-th sensor in the BMSC problem.
In particular, following the datasheet [22], for each index
$j = 1, \ldots, m - 1$, the calibration parameters $f_{1,j}$, $f_{2,j}$, and
$f_{3,j}$—defined in Eq. (1)—of the uncalibrated mobile sen-
sors are randomly set according to a Gaussian distribution
centered around 0.9, 13, and -2.5, respectively. We then
get a $31 \times 100$ theoretical observation matrix for which we
randomly keep $k + l$ samples, where $k$ (respectively, $l$) is
the number of calibrated (respectively, uncalibrated) sensor
samples—with $k \ll l$—hence providing the irregular sam-
pling over the scene. Lastly, Gaussian noise realizations may
be added to the observed uncalibrated sensor data.

In this section, we aim to investigate the performance
of the number of rendezvous between calibrated and uncali-
brated sensors, the number of missing entries in $X$, and the
influence of the input SNR to the BMSC performance. For
each test condition—i.e., one number of rendezvous, one pro-
portion of missing entries, or one input SNR—20 simulations
are performed. For each run, we randomly set the position
of the samples in $X$ in the three experiments and we generate
different noise realizations in the last one. The number $k$ of
calibrated sensor values in the $m$-th column of $X$ is set to 4
in all the tests. Except when we make these values vary, the
proportion of uncalibrated sensors to have rendezvous with
 calibrated ones, and the proportion of missing entries in $X$ are
set to 30 and 80 %, respectively. In addition to the noiseless
case, the input SNR varies from 10 to 80 dB. Lastly, the runs
are stopped after 5e5 iterations. The estimation error of each
calibration parameter is measured by the Root Mean Square
Error (RMSE) computed between a row of true parameters in
$F$ and the corresponding row of reconstructed ones. Their av-
average provides a global RMSE for the considered simulation.
It should be noticed that state-of-the-art nonlinear calibra-
tion methods [14, 15] cannot be tested in this experimental
section, because they require the observed signal subspace
knowledge and/or some time synchronization of the sensors
measurements—which is a strong assumption not satisfied
in most crowdsensing applications, and not needed by the
proposed method.

Figure 2 shows the plots of the different experiments re-
alized in this paper. The blue (respectively, red) line cor-
responds to the median RMSE obtained with our proposed
method (respectively, the matrix-completion-based approach
discussed in Subsection 3.3) applied on 20 independent tests
while the blue (respectively, red) area shows the envelope
of the RMSEs obtained with our proposed method (respect-
ively, the matrix-completion-based approach). The left plot
shows the influence of the proportion $\rho$ of missing values—
ranging from 0.1 to 0.9—on the obtained RMSE. Our pro-
posed method always outperforms the naive approach. Inter-
estingly, the median RMSE is very stable with respect to $\rho$:
it slightly increases when $\rho = 90$ % only. Lastly, it seems
strange that the low part of the envelope is almost decreasing
when $\rho$ increases (until 70 %). In fact, in several simu-
lations, the semi-NMF method did not converge to the op-
timal solution within 5e5 iterations, which should be solved
by running more iterations. The central plot shows the influ-
ence of the number rendezvous between calibrated and un-
calibrated sensors. Again, the proposed method outperforms
the naive approach and provides stable median RMSEs over
the rendezvous proportion. Strangely, the RMSE is not that
high when there is no rendezvous. In fact, the factorization
is obtained up to a scale ambiguity and, since the initializa-
tion is not too far from the true values, the RMSE remains
quite low. The right plot shows the influence of the input SNR
on the performance. The proposed approach outperforms
the naive method if the input SNR is above 35 dB. It should
provide a better performance at lower input SNRs with more data
points, e.g., by considering several scenes and by stacking the
data matrices into a unique bigger one.

5. CONCLUSION

In this paper, we proposed an informed nonlinear semi-NMF
method for mobile sensor calibration. The proposed approach
is using a Vandermonde structure in one factor and is able
to process the factorization with missing data. The proposed
method was tested on a crowdsensing-like simulation and was
found to provide a stable median performance over the miss-
ning value proportion and the number of rendezvous between
calibrated and uncalibrated sensors. In future work, we aim to
speed up the proposed approach, e.g., by estimating optimal
gradient step-sizes. Moreover, we aim to consider other non-
linear calibration functions, e.g., to deal with sensor response
drift or saturation.
6. REFERENCES


