A decentralised linear programming approach to energy-efficient event detection

Kun Yuan and Qing Ling*

Department of Automation,
University of Science and Technology of China,
Hefei, Anhui 230027, China
Email: kunyuan@mail.ustc.edu.cn
Email: qingling@mail.ustc.edu.cn
*Corresponding author

Zhi Tian

Department of Electrical and Computer Engineering,
Michigan Technological University,
Houghton, Michigan 49931, USA
Email: ztian@mtu.edu

Abstract: This paper addresses the problem of decentralised event detection in a large-scale wireless sensor network (WSN), where without the coordination of any fusion centre, sensor nodes autonomously make decisions through information exchange with their neighbours. We formulate the event detection problem as a linear program, and propose a heuristic decentralised linear programming (DLP) approach to solve it. Compared to the existing algorithms which solve decentralised optimisation problems, the DLP algorithm requires low communication cost per iteration and shows fast convergence. Convergence property of the DLP algorithm is theoretically analysed, and then validated through simulation results. We also implement the DLP algorithm in the structural health monitoring application and demonstrate its effectiveness in decentralised event detection.

Keywords: WSN; wireless sensor network; decentralised optimisation; event detection; linear programming.

Reference to this paper should be made as follows: Yuan, K., Ling Q. and Tian, Z. (2015) 'A decentralised linear programming approach to energy-efficient event detection', *Int. J. Sensor Networks*, Vol. 17, No. 1, pp.52–62.

Biographical notes: Kun Yuan received the BE in Telecommunication Engineering from Xidian University in 2011. He is now pursuing his MS in Control Theory and Control Engineering at the Department of Automation, University of Science and Technology of China. His current research focuses on decentralised optimisation of networked multi-agent systems.

Qing Ling received the BE in Automation and the PhD in Control Theory and Control Engineering from University of Science and Technology of China in 2001 and 2006, respectively. From 2006 to 2009, he was a postdoctoral research fellow with the Department of Electrical and Computer Engineering, Michigan Technological University. Since September 2009, he has been an Associate Professor with the Department of Automation, University of Science and Technology of China. His current research focuses on decentralised optimisation of networked multi-agent systems.

Zhi Tian received the BE in Automation from the University of Science and Technology of China in 1994, the MS and PhD in Electrical Engineering from George Mason University in 1998 and 2000, respectively. Since August 2000, she has been with the Department of Electrical and Computer Engineering, Michigan Technological University, where she is currently a Professor. Her research interests are in the areas of signal processing for wireless communications, estimation and detection theory. She served as Associate Editors for IEEE Transactions on Wireless Communications and IEEE Transactions on Signal Processing. She received a CAREER award in 2003 from the US National Science Foundation.

1 Introduction

Event detection is one of the most important applications of wireless sensor networks (WSNs) (Ling et al., 2010). Typical event detection tasks include detecting nuclear radioactive sources (Sundaresan et al., 2007), monitoring structural health conditions (Lynch, 2007), discovering presence of contaminants (Sun and Coyle, 2010), to name a few. In a traditional event detection WSN, data collection and signal processing are separated. A fusion centre collects sensory data from distributed wireless sensor nodes, and then processes the sensory data to accomplish the event detection task. Such a *centralised* infrastructure suffers from sharp degradations of energy efficiency and robustness when the network scale increases (Liu et al., 2011).

To address this issue, *decentralised* algorithms in WSNs have attracted considerable research interest in recent years (Rabbat and Nowak, 2006; Nedic and Ozdaglar, 2009; Predd et al., 2007; Zhu et al., 2007; Schizas et al., 2008; Bazerque and Giannakis, 2010; Ling and Tian, 2010; Jakovetic et al., 2011). In the absence of any fusion centre, sensor nodes autonomously exchange information with their neighbours and collaboratively detect the events. This decentralised infrastructure, compared to the traditional centralised one, avoids collecting sensory data to a fusion centre while exploits collaboration of neighbouring sensor nodes to realise in-network signal processing. Therefore, the network can achieve lightweight communication and improved energy efficiency. In addition, avoiding centralised collection of sensory data also strengthens network robustness since the network no longer relies on several critical sensor nodes to relay sensory data. For detailed discussions on the benefits of decentralised algorithms in WSNs, readers are referred to Predd et al. (2007).

This paper elaborates on developing an energy-efficient and robust decentralised algorithm, which scales with network size, for the event detection application of WSNs. Specifically, we confine the event sources to be at a set of candidate positions (here we use sensor node positions as candidate positions), and model the event detection problem as recovering a decision vector which represents the event magnitudes at the candidate positions. Though the existing decentralised in-network signal processing techniques, such as *consensus optimisation*, are applicable to this recovery problem, they often lead to high communication cost (Schizas et al., 2008; Bazerque and Giannakis, 2010; Ling and Tian, 2010; Jakovetic et al., 2011). For example, in the consensus optimisation formulation, each sensor node holds a local copy of the signal to recover, and the local copies of neighbouring sensor nodes consent to the same value. Decentralised iterative algorithms which solve the consensus optimisation problem require sensor nodes to exchange their current local copies with their neighbours; therefore, for each sensor node, the communication cost per iteration is proportional to the size of its local copy (Schizas et al., 2008; Bazerque and Giannakis, 2010; Ling and Tian, 2010; Jakovetic et al., 2011). Since the size of each local copy is equal to the number of sensor nodes in our model, the communication cost per sensor node increases at least linearly with the network size, which is unacceptable for large-scale WSNs.

This paper develops a decentralised linear programming (DLP) algorithm to solve the event detection problem in an energy-efficient manner. Different from the consensus optimisation technique, we set an alternative objective that is simple yet effective. Sensor nodes no longer optimise and consent on the entire decision vector as in the consensus optimisation approach; in contrast, we let each sensor node optimise its own decision variable corresponding to a scalar element in the decision vector. Indeed, each sensor node solves a simple linear program through limited communication with its neighbouring sensor nodes. This strategy works since the sensory data in the event detection application often exhibits local correlation other than global correlation; that is, an occurring event only influences the measurements of some nearby sensor nodes, but has little effect on faraway sensory readings. Therefore, recovering the whole decision vector at each sensor node and letting the whole network reach consensus are both unnecessary. Theoretically, we prove that under mild conditions, the DLP algorithm approximately solves the original event detection problem. We further show that for a certain scenario, the DLP algorithm converges to the exact solution; this scenario appears in applications such as structural health monitoring, etc. Compared to the alternating direction method (ADMM) algorithm which solves the consensus optimisation problem, the proposed DLP algorithm shows much faster convergence and lower communication cost through numerical simulation.

This paper is organised as follows. The event detection problem is formulated in Section 2. Section 3 proposes the DLP algorithm and Section 4 theoretically analyses its convergence property. Numerical simulation and application in structural health monitoring are demonstrated in Sections 5 and 6, respectively. Section 7 concludes the paper.

2 Problem formulation

Suppose that a large-scale WSN is densely deployed in a two-dimensional sensing field. The network has a set of L sensor nodes, denoted as $\mathcal{L} = \{v_i\}_{i=1}^{L}$. Sensor nodes have a common communication range r_C. Each sensor node can and only can communicate with its one-hop neighbours within the communication range, and no multi-hop communication is permitted to avoid excessive coordination efforts. Given the communication range r_C, the network is bidirectionally connected.

The event detection task is performed periodically. Within each sampling period, multiple events may occur in the sensing field. Each event has an influence on part of the sensing field, centring around the source of the event; influences of multiple events are then superposed on the whole sensing field. After collecting the sensory data, distributed sensor nodes cooperate to localise the sources and estimate the amplitudes of the events.

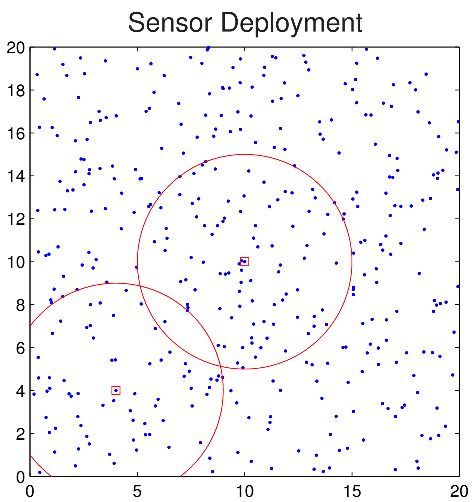
To make the event detection problem tractable, we assume that there are a set of *candidate positions* for the event sources. Typical choices of candidate positions include virtual grid points in the sensing field (Bazerque and Giannakis, 2010) and positions of sensor nodes (Ling and Tian, 2010).

Setting the virtual grid points as the candidate positions makes the resolution of event detection adjustable; however, in decentralised algorithm design, this setting requires each sensor node to detect all events in the whole sensing field, and hence brings high communication cost (Bazerque and Giannakis, 2010). Contrarily, by choosing the positions of the sensor nodes as the candidate positions, the resolution of event detection is directly decided by the density of the sensor nodes. When the WSN nodes are densely deployed, this setting makes sense. More importantly, this setting simplifies decentralised algorithm design since each sensor node only needs to detect whether there is an event occurring at its own position, as we shall see below. Therefore, we assume that

(A1) *Events occur only at the positions of some sensor nodes. When the source of one event coincides with the position of v_j, we denote the amplitude of the event by a scalar $c_j > 0$; otherwise $c_j = 0$.*

Based on (**A1**), we can formulate the event detection problem as recovering a nonnegative vector $\mathbf{c} = [c_1, ..., c_L]^T$ from the sensory data. Figure 1 shows the case where sensor nodes are deployed randomly in the sensing field and events occur at sensor node points.

Figure 1 Blue points are 400 sensor nodes deployed randomly in the sensing field, red squares are events, and red circles are events' influence areas (see online version for colours)



Next, we notice that in many event detection applications, one event only has a *limited influence* on its nearby region in the sensing field. For example, in nuclear radioactive detection, the influence of a nuclear source decreases polynomially in distance. Similar distance-dependent influences can be observed from events such as fire sources and structural damages. By carefully setting the communication range r_C, we can confine the influence of an event occurring at the position of one sensor node to the range of the sensor node's one-hop neighbours; more specifically,

(**A2**) *Each event only influences its nearby region in the sensing field. The influence of a unit-amplitude event occurring at the position of v_j on the sensor node v_i is f_{ij}, and $f_{ij} = f_{ji}$. We suppose that $f_{ij} = 0$ if $d_{ij} \geq r_C$, where d_{ij} is the distance between sensor nodes v_i and v_j.*

The assumption **(A2)** enables design of simple decentralised algorithms, since it eliminates the interrelationship between multi-hop sensor nodes and avoids the necessity of multi-hop communications.

Finally, we make a commonly-used assumption that the influences of the events are linearly superposed. Combining with **(A2)**, we have

(A3) *The measurement of one sensor node is the superposition of the influences of all events plus random noise. Since $f_{ij} = 0$ for $d_{ij} \geq r_C$, the measurement b_i of v_i can be written as $b_i = \sum_{v_j \in \mathcal{N}_i \cup v_i} f_{ij} c_j + e_i$, where e_i is random noise, and $\mathcal{N}_i$ denotes the one-hop neighbours of v_i.*

We now introduce a key observation that the vector $\mathbf{c}$, which we want to recover, is sparse. That is, the number of nonzero elements in $\mathbf{c}$ is much smaller than the vector size L. This prior knowledge holds since event detection for a large-scale WSN is meaningless if the sensing field is full of event sources. Nevertheless, we do allow the influences of these sparse events to span over the whole sensing field. Hence, reconstruction of $\mathbf{c}$ boils down to minimising a sparsity-imposing metric $||\mathbf{c}||_1$ that is the ℓ_1 norm of $\mathbf{c}$ (Donoho, 2006), subject to measurement constraints. From **(A3)**, two similar formulations arise, one is a linear program that postulates a bound θ of measurement noise at each sensor nodes:

$$\begin{aligned} \min_{\mathbf{c}} \quad & ||\mathbf{c}||_1 \\ \text{s.t.} \quad & |b_i - \textstyle\sum_{v_j \in \mathcal{N}_i \cup v_i} f_{ij} c_j| \leq \theta, \quad \forall v_i \in \mathcal{L} \\ & c_i \geq 0, \quad \forall v_i \in \mathcal{L}. \end{aligned} \tag{1}$$

And the other is a second-order cone program that confines the total energy of the measurement noise to be lower than ϵ:

$$\begin{aligned} \min_{\mathbf{c}} \quad & ||\mathbf{c}||_1 \\ s.t. \quad & \textstyle\sum_{v_i \in \mathcal{L}} (b_i - \sum_{v_j \in \mathcal{N}_i \cup v_i} f_{ij} c_j)^2 \leq \epsilon \\ & c_i \geq 0, \quad \forall v_i \in \mathcal{L}. \end{aligned} \tag{2}$$

Both equations (1) and (2) incorporate the prior knowledge of sparse events, which is important in alleviating the undesired false alarm rates of otherwise non-sparse solutions produced by general signal recovery approaches such as thresholding or least squares. In this paper, we focus on the linear program equation (1) and devote to designing an efficient algorithm as well as analysing its recovery performance.

3 Decentralised algorithms

Indeed, we can reformulate the event detection problem (1) as a consensus optimisation problem via introducing local copies of $\mathbf{c}$ and imposing equality constraints on the local copies of neighbouring sensor nodes. This consensus optimisation problem can be solved based on the alternating direction method of multipliers (ADMM) (Bertsekas and Tsitsiklis, 1997). For details of the ADMM-based decentralised algorithm, readers are referred to Bazerque and Giannakis (2010) and Ling and Tian (2010). However, the resulting decentralised algorithm requires each sensor node to share its current local copy of $\mathbf{c}$ with its neighbouring sensor nodes. Since in our model, the size of $\mathbf{c}$ is equal to the number of sensor nodes, this approach does not scale when the network size increases. Further, the consensus optimisation scheme requires that all sensor nodes reach a consensus, which often brings slow convergence rate for a large-scale WSN.

To reduce the communication cost per iteration as well as accelerate the algorithm, this paper proposes a heuristic DLP algorithm to solve equations (1). Specifically, sensor node v_i solves the following linear program:

$$\begin{aligned} \min_{c_i} \quad & c_i \\ \text{s.t.} \quad & |b_i - f_{ii}c_i - \textstyle\sum_{v_j \in \mathcal{N}_i} f_{ij}c_j| \leq \theta \\ & c_i \geq 0. \end{aligned} \tag{3}$$

Indeed, equation (3) is corresponding to part of the objective function and part of the constraints of equation (1).

Solution to equation (3) is $c_i = [b_i - \theta - \sum_{j \in \mathcal{N}_i} f_{ij}c_j]^+/f_{ii}$. Here $[\cdot]^+$ denotes the projection to $\max\{\cdot, 0\}$. Therefore we have the following heuristic DLP algorithm as in Algorithm 1.

Algorithm 1 Decentralized Linear Programming (DLP) Algorithm for sensor v_i

---

Require: Event influence set $\mathcal{N}_i$, coefficient vector $\mathbf{F}_i = [f_{i1},\ f_{i2},\ \cdots,\ f_{iL}]^T$, measurement b_i and predefined threshold θ

1: Initialize $c_i = 0$;
2: **for** $t = 0, 1, 2, \ldots$, sensor v_i **do**
3: Transmit $c_i(t)$ to, and receive $c_j(t)$ from $j \in \mathcal{N}_i$;
4: Update $c_i(t+1)$ according to $c_i = [b_i - \theta - \sum_{j \in \mathcal{N}_i} f_{ij}c_j]^+/f_{ii}$
5: **end for**
6: Returns $c_i(t+1)$.

Compared with the ADMM-based decentralised algorithm, the DLP algorithm just needs to exchange its own decision variables c_i, other than the local copy of $\mathbf{c}$ hold by itself with its neighbours. This way, the communication cost per iteration can be greatly reduced. Further, as we will demonstrate with numerical simulation, the DLP algorithm converges much faster than the ADMM-based decentralised algorithm. The two properties make the DLP algorithm very energy-efficient in the event detection application.

4 Convergence property of DLP

In this section, we elaborate on the convergence property of the proposed DLP algorithm. First, we prove that under mild conditions, the DLP algorithm solves an optimisation problem which is similar to the original linear program (1). Second, we further show that for a certain scenario, the DLP algorithm exactly converges to the optimal solution of equation (1); this scenario appears in applications such as structural health monitoring, etc.

Before going to the theoretical analysis, we rewrite equation (1) to its equivalent matrix form as

$$\begin{aligned} \min_{\mathbf{c}} \quad & \mathbf{1}^T\mathbf{c} \\ \text{s.t.} \quad & \mathbf{Fc} - \mathbf{b} \geq -\theta\mathbf{1} \\ & \mathbf{Fc} - \mathbf{b} \leq \theta\mathbf{1} \\ & \mathbf{c} \geq \mathbf{0}. \end{aligned} \tag{4}$$

Here $\mathbf{1} = [1, 1, ..., 1]^T$ is an $L \times 1$ vector, $\mathbf{b} = [b_1, b_2, ..., b_L]^T$ contains sensory measurements, and

$$\mathbf{F} = \begin{pmatrix} f_{11} & \cdots & f_{1L} \\ \vdots & \ddots & \vdots \\ f_{L1} & \cdots & f_{LL} \end{pmatrix}.$$

In the matrix $\mathbf{F}$, all diagonal elements are 1; an off-diagonal element $f_{ij} \geq 0$ and $f_{ij} = 0$ if $v_j \notin \mathcal{N}_i \cup v_i$. Since $f_{ij} = f_{ji}$, $\mathbf{F}$ is symmetric.

Proposition 1: *If a symmetric non-negative matrix* $\mathbf{F}$ *is strictly diagonally dominant and all of its diagonal elements are equal to 1, then the proposed DLP algorithm converges to the optimal solution of a linear programming problem*

$$\begin{aligned} \min_{\mathbf{c}} \quad & \mathbf{1}^T\mathbf{Fc} \\ \text{s.t.} \quad & \mathbf{Fc} - \mathbf{b} \geq -\theta\mathbf{1} \\ & \mathbf{c} \geq \mathbf{0}. \end{aligned} \tag{5}$$

Proof: It is proved in Mangasarian (1976) that if a nonnegative matrix $\mathbf{F}$ is strictly diagonally dominant, then the optimal solution of linear program (5) is equivalent to that of the following linear complementarity problem:

$$\mathbf{c}^T(\mathbf{Fc} - \mathbf{b} + \theta\mathbf{1}) = \mathbf{0},\ \text{s.t.}\ \mathbf{Fc} \geq \mathbf{b} - \theta\mathbf{1},\ \mathbf{c} \geq \mathbf{0}. \tag{6}$$

Let us rewrite (6) into a quadratic program

$$\begin{aligned} \min_{\mathbf{c}} \quad & \mathbf{c}^T(\mathbf{Fc} - \mathbf{b} + \theta\mathbf{1}) \\ \text{s.t.} \quad & \mathbf{Fc} - \mathbf{b} \geq -\theta\mathbf{1} \\ & \mathbf{c} \geq \mathbf{0}. \end{aligned} \tag{7}$$

As proved in Mangasarian (1991), if $\mathbf{F} = \mathbf{P} + \mathbf{Q}$ and $\mathbf{P} - \mathbf{Q}$ are both positive definite, and $\mathbf{F}$, $\mathbf{P}$, and $\mathbf{Q}$ are all symmetric, then the quadratic program (7) can be solved by a matrix splitting algorithm $\mathbf{c}(t+1) = [\mathbf{c}(t+1) - (\mathbf{Pc}(t+1) + \mathbf{Qc}(t) - \mathbf{b} + \theta\mathbf{1})]^+$.

Let $\mathbf{P} = \mathbf{I}$ be an $L \times L$ identity matrix, then $\mathbf{Q} = \mathbf{F} - \mathbf{I}$. Obviously $\mathbf{F}$, $\mathbf{P}$, and $\mathbf{Q}$ are all symmetric matrices. Note that $\mathbf{F}$ and $\mathbf{P} - \mathbf{Q} = 2\mathbf{I} - \mathbf{F}$ are both symmetric and strictly diagonally dominant, $\mathbf{F}$ and $\mathbf{P} - \mathbf{Q}$ are both positive definite. The corresponding matrix splitting algorithm to (7) is hence $\mathbf{c}(t+1) = [\mathbf{b} - \theta\mathbf{1} - (\mathbf{F} - \mathbf{I})\mathbf{c}(t)]^+$, which is equivalent to the DLP algorithm. Therefore, the proposed DLP algorithm converges to the optimal solution of equation (5). □

Remark 1: In Proposition 1, the only notable assumption is that $\mathbf{F}$ is strictly diagonally dominant. It means that the influence of an event decays fast enough with respect to distance. That is, if an event occurs at the position of sensor node v_i with unit magnitude $c_i = 1$, then its accumulated influence on other sensor nodes is less than 1, i.e., $\sum_{v_j \in \mathcal{N}_i} f_{ij} < 1$.

Proposition 1 shows that under mild conditions, the DLP algorithm solves equation (5), a linear programming problem similar to the original problem (4). The objective function of equation (5) is $\mathbf{1}^T \mathbf{F}\mathbf{c}$, which reweights the objective function of equation (4), i.e., $\mathbf{1}^T \mathbf{c}$. In addition, the solution of equation (5) is not necessarily satisfying the inequality constraint $\mathbf{F}\mathbf{c} - \mathbf{b} \leq \theta \mathbf{1}$. Note that Proposition 1 only gives sufficient conditions on the equivalence; the strictly diagonally dominant assumption can be relaxed, as we will demonstrate in the simulation results.

Next, we will give an important proposition which guarantees that the DLP algorithm exactly solves equation (4). This proposition shows that if the events' magnitudes are large enough (i.e., the signal-to-noise ratio is large enough), the DLP algorithm can converge to the optimal solution of equation (4) with a proper choice of the parameter θ.

Summing on the rows of the coefficient matrix $\mathbf{F}$, we denote $s_{\max}$ and $s_{\min}$ the maximum and minimum values, respectively. Besides, we denote $\mathbf{c}_0$ as the vector which represents real event occurrence in the sensing field. If $c_{0i} \neq 0$, then an event occurs at the position of sensor node v_i with magnitude c_{0i}. Moveover, we denote $\mathbf{n}$ as the noise involved in measurement $\mathbf{b}$. The following proposition proves when the DLP algorithm exactly solves the original problem (4).

Proposition 2: *Suppose that* $\mathbf{F}$ *is symmetric non-negative, strictly diagonally dominant, all of its diagonal elements are equal to 1, and* $\frac{s_{\min}-1}{s_{\max}-1} > \frac{1}{\beta} > 0$. *Then for any event occurring at the position of sensor node* v_i *with* $c_{0i} \geq (\beta + 1) \|\mathbf{n}\|_\infty$, *if we set* $\theta \geq \beta \|\mathbf{n}\|_\infty$, *the DLP algorithm exactly solves the original problem (4).*

Proof: First, in Appendix A.III, we prove that under the above conditions, equation (4) is equivalent to:

$$\begin{aligned} \min_{\mathbf{c}} \quad & \mathbf{1}^T \mathbf{c} \\ \text{s.t.} \quad & \mathbf{F}\mathbf{c} - \mathbf{b} \geq -\theta \mathbf{1} \\ & \mathbf{c} \geq \mathbf{0}. \end{aligned} \tag{8}$$

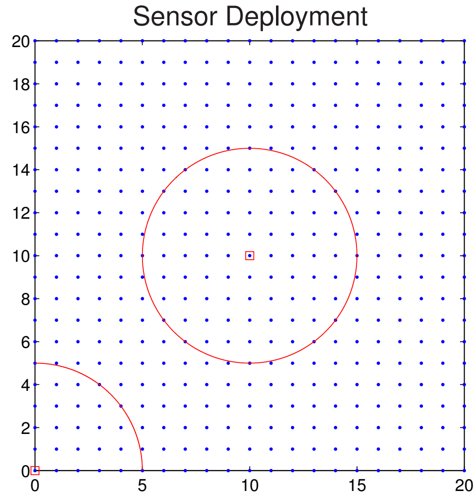
Second, in Proposition 1, we prove that if $\mathbf{F}$ is symmetric non-negative, strictly diagonally dominant, and all of its diagonal elements are equal to 1, then the DLP algorithm converges to the optimal solution of equation (5). Therefore, to prove Proposition 2, we only need to show the equivalence between equations (8) and (5), which is shown in Appendix A.IV. □

Remark 2: Proposition 2 indicates that given $s_{\max}$ and $s_{\min}$, if the signal-to-noise ratio is larger than a certain threshold (for any event occurring at v_i, $c_{0i} \geq (\beta + 1) \|\mathbf{n}\|_\infty$) and the parameter θ is properly chosen ($\theta \geq \beta \|\mathbf{n}\|_\infty$), then the DLP algorithm exactly solves the original problem (4).

Remark 3: Proposition 2 shows that when $\theta > \beta\|\mathbf{n}\|_\infty$, the DLP converges to (4). Actually, $\beta\|\mathbf{n}\|_\infty$, as shown in Appendix A, is quite a conservative bound for θ. Simulation results show that for $\|\mathbf{n}\|_\infty < \theta < \beta\|\mathbf{n}\|_\infty$, DLP may also be able to converge to equation (4).

Proposition 2 gives a guideline of adjusting β based on the values of $s_{\max}$ and $s_{\min}$. When sensor nodes are randomly deployed, we can estimate $s_{\max}$ and $s_{\min}$ via the statistical property of $\mathbf{F}$. For applications which allow us deploy sensor nodes in a regular manner, $s_{\max}$ and $s_{\min}$ can be exactly determined. Next we show a special example where sensor nodes are placed at virtual grid points in the square sensing field, as shown in Figure 2. In the structure health monitoring application, as we will show in Section 4, sensor nodes can often be deployed in this manner.

Figure 2 Blue points are 400 sensor nodes deployed at virtual grid points, red solid squares are events, and red circles are events' influence ranges (see online version for colours)



Corrolary 3: *If sensors are deployed as in Figure 2, Then for any event occurring at the position of sensor node* v_i *with* $c_{0i} \geq 5 \|\mathbf{n}\|_\infty$, *if we set* $\theta \geq 4 \|\mathbf{n}\|_\infty$, *the DLP algorithm exactly solves equation (4).*

Proof: If sensors are deployed as in Figure 2, it is easy to deduce that $\frac{s_{\min}-1}{s_{\max}-1} > \frac{1}{4}$. And from Proposition 2, Corollary 3 is established. □

5 Simulation results

In our simulation, we divide the sensing field into an $N \times N$ lattice, and deploy $L = N^2$ sensor nodes at the grid points. The distance between two neighbouring sensor nodes is r, such that the lattice grid can be represented by sensor positions $\{(xr, yr) : x, y = 1, 2, ..., N\}$. Suppose that events occur at several grid points. We use the following basis function to approximate the influence of unit-magnitude event:

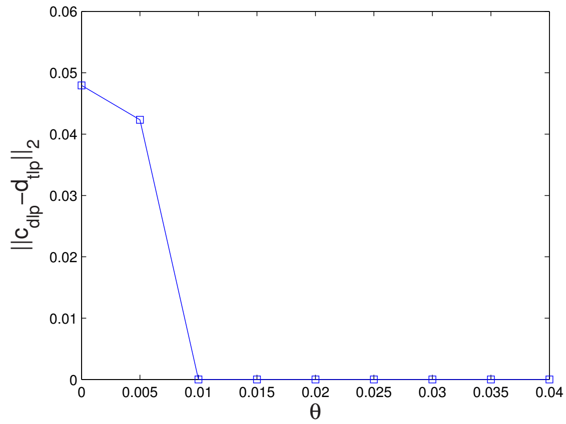
$$f_{ij} = f_j(v_i) = \begin{cases} e^{-\frac{d_{ij}^2}{\sigma^2}} & d_{ij} < r_E \\ 0 & d_{ij} \geq r_E \end{cases}. \tag{9}$$

Here we let $r_C = r_E$. In addition, sensory measurements are polluted by uniformly random noise, ranging from -0.01 to 0.01.

First we check the convergence property of the DLP algorithm. Suppose four events with magnitude $0.25, 0.5, 0.75$ and 1 (satisfying $c_{0i} \geq 5\|\mathbf{n}\|_\infty$) occur at $(3r, 5r), (4r, 2r), (6r, 6r)$ and $(8r, 7r)$, respectively. Meanwhile, σ^2 is set as 0.5 and $r_E = 0.01$ to keep $\mathbf{F}$ being strictly diagonally dominant. Denote $\mathbf{c}_0$ as the true event vector, $\mathbf{c}_{tlp}$ as the optimal solution of problem (4) and $\mathbf{c}_{dlp}$ as the fixed point of DLP. $\|\mathbf{c}_{tlp} - \mathbf{c}_{dlp}\|_2$ denotes the gap between the fixed point of DLP and optimal solution of (4). During the simulation, θ varies from 0 to 0.04 while $\|\mathbf{n}\|_\infty$ is fixed at 0.01.

Figure 3 depicts that when $\theta > 0.01$, DLP converges to the optimal solution of equation (4). Actually, from the deduction of Proposition 2 we know it is quite conservative to choose $\theta > 4\|\mathbf{n}\|_\infty$. DLP performs much better than our theoretical analysis.

Figure 3 DLP converges to the optimal solution of equation (4) when $\theta > 0.01$ (see online version for colours)



Next we compare the convergence between DLP and ADMM. Suppose two events, with magnitudes 1 and 0.5, occur at $(3r, 5r)$ and $(5r, 5r)$. During the simulation, $\sigma^2 = 0.5$, $r_E = 0.01$, $\|\mathbf{n}\|_\infty = 0.01$ and $\theta = 0.05$. Figure 4 shows the convergence behaviour of the two algorithms. These two decentralised algorithms both provide accurate locations and converge to the solution of model (4). However, the ADMM algorithm converges after 90 iterations, while the DLP algorithm converges after only 4 iterations. With less communication cost per iteration and much faster convergence rate, DLP greatly improves energy-efficiency of the network.

Finally we will show DLP can also perform well in some cases where $\mathbf{F}$ is not diagonally dominant. Note that we can adjust σ^2 to vary the diagonal dominance of $\mathbf{F}$ and it is easy to verify that when $\sigma^2 > 0.65$ and $r_E = 0.01$, $\mathbf{F}$ is not diagonally dominant. Figure 5 shows that when $\sigma^2 \leq 1.2$, DLP will always converge exactly to the optimal solution of problem (4), but after that, DLP can hardly converge.

6 Structural health monitoring

This section demonstrates the application of the DLP algorithm in a structure health monitoring (SHM) network. SHM refers to the process of damage detection for civil, aerospace and mechanical engineering systems (Lynch, 2007). The application of WSNs in SHM has attracted much attention in recent years. By collecting the structure health information through distributed and unattended sensor nodes, damages or anomalies in a structure can be localised and the severities of damages or anomalies can be identified.

Figure 4 DLP vs. ADMM (see online version for colours)

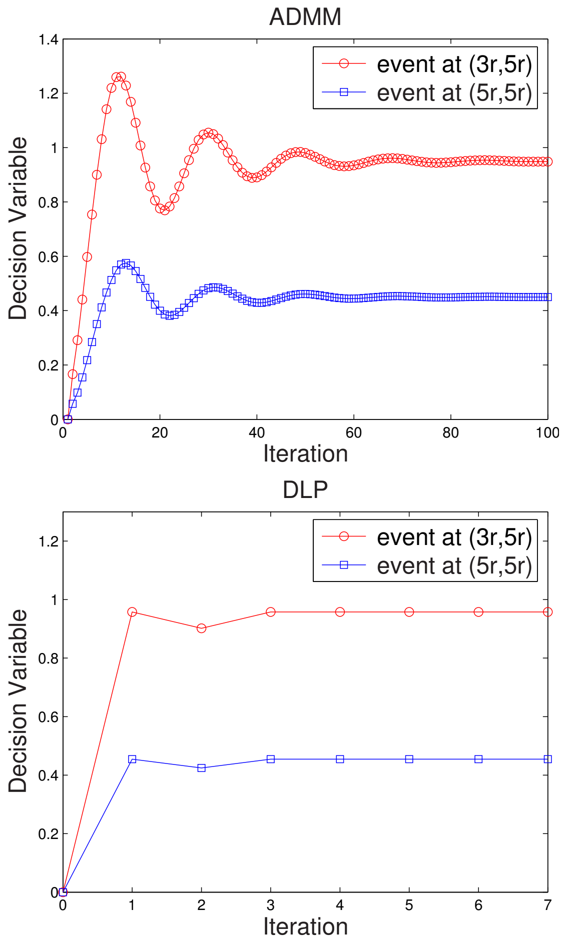
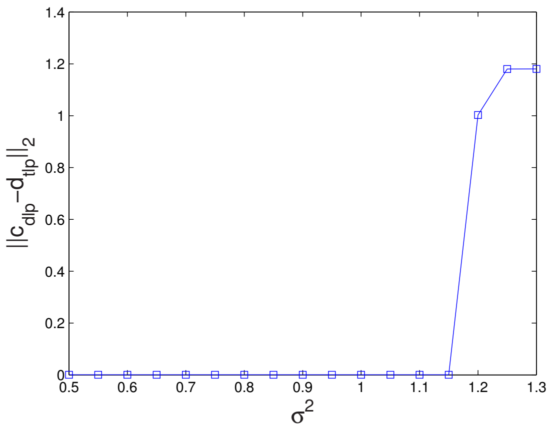


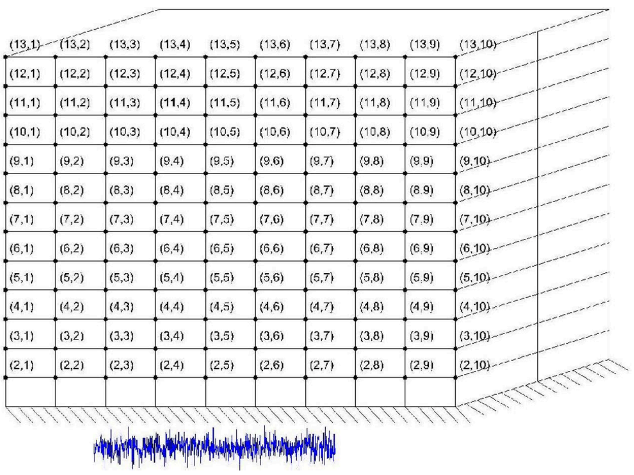
Figure 5 DLP can still converge to the optimal solution of equation (4) even if $\mathbf{F}$ is not strictly diagonally dominant. Note DLP cannot converge when $\sigma^2 \geq 1.2$, where we force it to stop after 200 iterations (see online version for colours)



The decentralised SHM system works with three stages (Ling et al., 2009; Ling and Tian, 2009). Firstly in a modelling stage, sensor nodes collect baseline responses from a undamaged structure. Secondly in a statistical pattern recognition stage, each sensor node collects operational responses and achieves a damage sensitive coefficient by comparing with baseline responses. Thirdly in a decentralised decision-making stage, the damage sensitive coefficients are used to estimate a vector of damage severity coefficients, which denotes the positions and severities of damages. Since each damage in a structure only influences the nearby sensor nodes and the vector of damage severity coefficients is often sparse. By restricting the locations of damages on sensor node points, we can establish an event detection model as in equation (1).

Consider the example in Ling and Tian (2009). Suppose that there is a steel frame structure with 12 stories and 9 bays, simplified as a two-dimensional model, as illustrated in Figure 6. A grid network of 120 sensor nodes is deployed at the joint points. The width of a bay is 24 feet and the height of a floor is 14 feet. Ambient vibrations are imposed to the foundation with Gaussian white noise. Responses of the structure are analysed by the finite element software OpenSees (Pacific Earthquake Engineering Research Center, 1999). By reducing 72% stiffness for the column between sensors (7,5) (located in 7th floor, 5th bay) and (8,5) (located in 8th floor, 5th bay), two damages are introduced to the structure. A typical spatial distribution of the damage sensitive coefficients, namely $\mathbf{b}$, is shown in Figure 7.

Figure 6 Two-dimensional model of a steel frame structure with 12 stories and 9 bays. Gaussian random white noise is imposed to the foundation to simulate ambient vibrations (see online version for colours)



The measurement matrix $\mathbf{F}$ is decided through simulating the structure. Denoting the positions of v_i and v_j as (m_i, n_i) and (m_j, n_j), respectively, simulation results indicate that the measurement coefficient $f_{ij} \simeq \exp(-(m_i - m_j)^2 - (n_i - n_j)^2)$. The communication range r_C is set to be slightly larger than 24 feet, such that each sensor can communicate with 4 neighbouring sensors. The threshold $\theta = 0.03$ in the DLP algorithm. As depicted in Figure 8, even when the matrix $\mathbf{F}$ under these setting is not diagonally dominant, we can still successfully detect damages at (7,5) and (8,5) with the DLP algorithm. We can also observe that the DLP algorithm converges within 8 iterations, a tremendously fast convergence rate.

Figure 7 Spatial distribution of the damage sensitive coefficients after reducing stiffness for the column between (7,5) and (8,5) (see online version for colours)

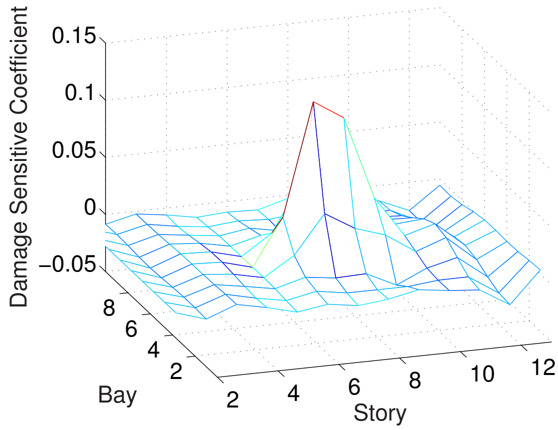
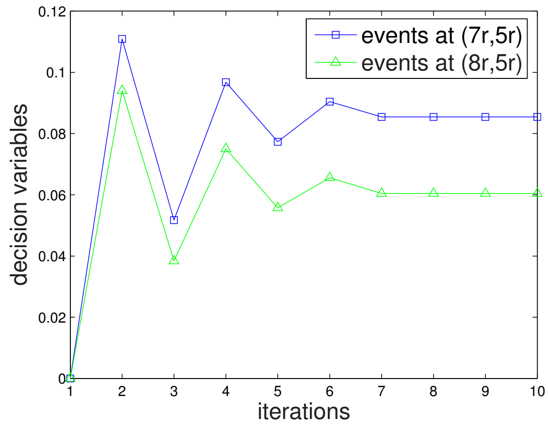


Figure 8 Convergence of the damage severity coefficients for sensors $(7, 5)$ and $(8, 5)$ when $\theta = 0.03$ (see online version for colours)



7 Conclusion

In this paper, we consider energy-efficient event detection with a WSN. We formulate the event detection problem as a linear program and design a heuristic DLP algorithm to solve it. The DLP algorithm is with light-weight communication cost per iteration and fast convergence rate. The convergence property of the DLP algorithm is analysed theoretically. Numerical simulation in both the synthetic examples and the SHM application shows that the DLP algorithm outperforms the ADMM-based consensus optimisation method.

References

Bazerque, J. and Giannakis, G. (2010) 'Distributed spectrum sensing for cognitive radio networks by exploiting sparsity', *IEEE Transactions on Signal Processing*, Vol. 58, pp.1847–1862.

Bertsekas, D. and Tsitsiklis, J. (1997) *Parallel and Distributed Computation: Numerical Methods*, 2nd ed., Athena Scientific.

Dantzig, G. (1963) *Linear Programming and Extensions*, Princeton University Press, Princeton, New Jersey, USA.

Donoho, D. (2006) 'Compressed sensing', *IEEE Transactions on Signal Processing*, Vol. 52, pp.1289–1306.

Jakovetic, D., Xavier, J. and Moura, J. (2011) 'Cooperative convex optimization in networked systems: augmented Lagrangian algorithms with direct gossip communication', *IEEE Transactions on Signal Processing*, Vol. 59, pp.3889–3902.

Ling, Q. and Tian, Z. (2009) 'A decentralized Gauss-Seidel approach for in-network sparse signal recovery', *Information Fusion, 2009. FUSION'09. 12th International Conference on*, IEEE, Seattle, USA, pp.380–387.

Ling, Q. and Tian, Z. (2010) 'Decentralized sparse signal recovery for compressive sleeping wireless sensor networks', *IEEE Transactions on Signal Processing*, Vol. 58, pp.3816–3827.

Ling, Q., Fanzi, Z. and Tian, Z. (2010) 'Energy-efficient decentralized event detection in large-scale wireless sensor networks', *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, IEEE, Dallas, USA, pp.3386–3389.

Ling, Q., Tian, Z., Yin, Y. and Li, Y. (2009) 'Localized structural health monitoring using energy-efficient wireless sensor networks', *IEEE Sensor Journal*, Vol. 9, pp.1596–1604.

Liu, Y., He, Y., Li, M., Wang, J., Liu, K., Mo, L., Dong, W., Yang, Z., Xi, M., Zhao, J. and Li, X. (2011) 'Does wireless sensor network scale? A measurement study on GreenOrbs', *INFOCOM, 2011 Proceedings IEEE*, IEEE, Shanghai, China, pp.873–881.

Lynch, J. (2007) 'An overview of wireless structural health monitoring for civil structures', *Philosophical Transactions of the Royal Society A*, Vol. 365, pp.345–372.

Mangasarian, O. (1976) 'Solution of linear complementarity problems by linear programming', *Numerical Analysis*, Vol. 506, pp.166–175.

Mangasarian, O. (1991) 'Convergence of iterates of an inexact matrix splitting algorithm for the symmetric monotone linear complementarity problem', *SIAM Journal of Optimization*, Vol. 1, pp.114–112.

Nedic, A. and Ozdaglar, A. (2009) 'Distributed subgradient methods for multi-agent optimization', *IEEE Transactions on Automatic Control*, Vol. 54, pp.48–61.

Pacific Earthquake Engineering Research Center (1999) *OpenSees: Open System for Earthquake Engineering Simulation*.

Predd, J., Kulkarni, S. and Poor, V. (2007) 'Distributed learning in wireless sensor networks', *IEEE Signal Processing Magazine*, Vol. 24, pp.56–69.

Rabbat, M. and Nowak, R. (2006) 'Quantized incremental algorithms for distributed optimization', *IEEE Journal on Selected Areas in Communications*, Vol. 23, pp.798–808.

Schizas, I., Ribeiro, A. and Giannakis, G. (2008) 'Consensus in ad hoc WSNs with noisy links – Part I: distributed estimation of deterministic signals', *IEEE Transactions on Signal Processing*, Vol. 56, pp.350–364.

Sun, X. and Coyle, E. (2010) 'Low-complexity algorithms for event detection in wireless sensor networks', *IEEE Journal on Selected Areas in Communications*, Vol. 28, pp.1138–1148.

Sundaresan, A., Varshney, P.K. and Rao, N.S.V. (2007) 'Distributed detection of a nuclear radioactive source using fusion of correlated decisions', *Information Fusion, 2007 10th International Conference on*, IEEE, Quebec, Canada, pp.1–7.

Zhu, Y., Liu, Y., Ni, L.M. and Zhang, Z. (2007) 'Low-power distributed event detection in wireless sensor networks', *INFOCOM 2007. 26th IEEE International Conference on Computer Communications. IEEE*, IEEE, Anchorage, USA, pp.2401–2405.

Appendix A

Notations: In this part upper curlicue letters are used for index set. By projecting the elements of the column vector $\mathbf{a}$, the rows and the columns of the matrix $\mathbf{A}$ on the index set $\mathcal{I}$, we achieve $\mathbf{a}_{\mathcal{I}}$, $\mathbf{A}_{(\mathcal{I},:)}$ and $\mathbf{A}_{(:,\mathcal{I})}$ respectively. Further, we always use $\mathbf{A}_{\mathcal{I}}$ in place of $\mathbf{A}_{(:,\mathcal{I})}$ for simplicity.

I. Preliminary: the simplex method

Introduce slack variables $c_{L+1}, c_{L+2}, ..., c_{2L}$ and rewrite equation (8) into

$$\begin{aligned} \max_{\mathbf{x}} \quad & \mathbf{e}^T\mathbf{x} \\ \text{s.t.} \quad & \mathbf{A}\mathbf{x} = \mathbf{d} \\ & \mathbf{x} \geq \mathbf{0}, \end{aligned} \tag{10}$$

where $\mathbf{e} = [-\mathbf{1}^T, \mathbf{0}^T]^T$ with $-\mathbf{1} = [-1, \cdots, -1]^T$ and $\mathbf{0} = [0, \cdots, 0]^T$ both being $L \times 1$ vectors, $\mathbf{d} = [-b_1 + \theta, \cdots, -b_L + \theta]^T$, $\mathbf{x} = [c_1, \cdots, \ c_L, c_{L+1}, \cdots, c_{2L}]^T$, and $\mathbf{A} = [-\mathbf{F}, \mathbf{I}]$.

Let $\mathcal{B}$ and $\mathcal{N}$ denote the index set of *basic variables* and *nonbasic variables*, respectively. Besides, define $\mathbf{B} = \mathbf{A}_{\mathcal{B}}$ and $\mathbf{N} = \mathbf{A}_{\mathcal{N}}$, then equation (10) can be transformed into:

$$\begin{aligned} \max_{\mathbf{x}_{\mathcal{B}}, \mathbf{x}_{\mathcal{N}}} \quad & \mathbf{e}_{\mathcal{B}}^T\mathbf{x}_{\mathcal{B}} + \mathbf{e}_{\mathcal{N}}^T\mathbf{x}_{\mathcal{N}} \\ \text{s.t.} \quad & \mathbf{B}\mathbf{x}_{\mathcal{B}} + \mathbf{N}\mathbf{x}_{\mathcal{N}} = \mathbf{d} \\ & \{\mathbf{x}_{\mathcal{B}}, \mathbf{x}_{\mathcal{N}}\} \geq \mathbf{0} \end{aligned} \tag{11}$$

Since $\mathbf{B}$ is nonsingular, equation (11) can be transformed into:

$$\begin{aligned} \max_{\mathbf{x}_{\mathcal{B}}, \mathbf{x}_{\mathcal{N}}} \quad & \mathbf{e}_{\mathcal{B}}^T\mathbf{B}^{-1}\mathbf{d} - ((\mathbf{B}^{-1}\mathbf{N})^T\mathbf{e}_{\mathcal{B}} - \mathbf{e}_{\mathcal{N}})^T\mathbf{x}_{\mathcal{N}} \\ \text{s.t.} \quad & \mathbf{x}_{\mathcal{B}} = \mathbf{B}^{-1}\mathbf{d} - \mathbf{B}^{-1}\mathbf{N}\mathbf{x}_{\mathcal{N}} \\ & \{\mathbf{x}_{\mathcal{B}}, \mathbf{x}_{\mathcal{N}}\} \geq \mathbf{0} \end{aligned} \tag{12}$$

Lemma 4: *Suppose* $\mathbf{x}^* \in \mathcal{R}^{2L}$. *If* $\mathbf{B}^{-1}\mathbf{d} \geq \mathbf{0}$ *and* $(\mathbf{B}^{-1}\mathbf{N})^T\mathbf{e}_{\mathcal{B}} - \mathbf{e}_{\mathcal{N}} \geq \mathbf{0}$, *let* $\mathbf{x}_{\mathcal{B}}^* = \mathbf{B}^{-1}\mathbf{d}$ *and* $\mathbf{x}_{\mathcal{N}}^* = 0$, *then* $\mathbf{x}^*$ *is optimal for problem (10).*

Proof: See Dantzig (1963). □

Define the index set $\mathcal{I} = \{1, 2, ..., 2L\}, \mathcal{I}_1 = \{1, 2, ..., L\}$ and $\mathcal{I}_2 = \{L+1, L+2, ..., 2L\}$, $\mathcal{B}_1 = \{j \in \mathcal{I}_1, j \in \mathcal{B}\}$, $\mathcal{B}_2 = \{j \in \mathcal{I}_2, j \in \mathcal{B}\}$, $\mathcal{N}_1 = \{j \in \mathcal{I}_1, j \in \mathcal{N}\}$ and $\mathcal{N}_2 = \{j \in \mathcal{I}_2, j \in \mathcal{N}\}$.

Lemma 5: *Suppose* x^* *is optimal for problem (10), then* $x_{\mathcal{I}_1}^*$ *is optimal for equation* (8). *If* $\mathbf{F}_{\mathcal{B}_1}\mathbf{x}_{\mathcal{B}_1}^* \leq \mathbf{b} + \theta\mathbf{1}$, *the solution of equation* (8) *equals that of equation* (4).

Proof: The first half is obvious, we now prove the second half. Note that the objective functions of the two problem are exactly the same, and hence we just need to prove $\mathbf{F}\mathbf{x}^*_{\mathcal{I}_1} \le \mathbf{b} + \theta\mathbf{1}$. Since $\mathbf{F}\mathbf{x}^*_{\mathcal{I}_1} = \mathbf{F}_{\mathcal{B}_1}\mathbf{x}^*_{\mathcal{B}_1} + \mathbf{F}_{\mathcal{N}_1}\mathbf{x}^*_{\mathcal{N}_1} = \mathbf{F}_{\mathcal{B}_1}\mathbf{x}^*_{\mathcal{B}_1} \le \mathbf{b} + \theta\mathbf{1}$, and hence the lemma is established. □

II. Preliminary: property of the matrix $\mathbf{F}$

Lemma 6: *Suppose that* $\mathbf{F}$ *is symmetric non-negative, strictly diagonally dominant, all of its diagonal elements are equal to 1. Then*

- $\mathbf{0} \le \mathbf{F}^{-1}\mathbf{1} \le \mathbf{1}$.
- *Denote* $s_{\max}$ *and* $s_{\min}$ *as the maximum and minimum row's sums of matrix* $\mathbf{F}$. *Given that* $\frac{s_{\min}-1}{s_{\max}-1} > \frac{1}{\beta} > 0$, *then if* $\theta \ge \beta\|\mathbf{n}\|_\infty$, *we have* $\mathbf{0} \le \mathbf{F}^{-1}(\theta\mathbf{1} - \mathbf{n}) \le \theta\mathbf{1} - \mathbf{n}$.

Proof: See Appendix B. □

III. Proof of the equivalence between equations (4) and (8)

Suppose that $\mathbf{c}_0$ is the true event vector. Define $\mathcal{Z} = \{i \in \mathcal{I}_1 : c_{0i} = 0\}$, $\bar{\mathcal{Z}} = \{i \in \mathcal{I}_1 : c_{0i} > 0\}$, $\mathcal{Z} + L = \{j : j = i + L, \forall i \in \mathcal{Z}\}$ and $\bar{\mathcal{Z}} + L = \{j : j = i + L, \forall i \in \bar{\mathcal{Z}}\}$. The train of thought to prove Proposition 2 is as follows:

Let $\mathcal{N} = \mathcal{Z} \cup (\bar{\mathcal{Z}} + L)$ and $\mathcal{B} = \bar{\mathcal{Z}} \cup (\mathcal{Z} + L)$. Meanwhile, define $\mathbf{B} = \mathbf{A}_{\mathcal{B}}$ and $\mathbf{N} = \mathbf{A}_{\mathcal{N}}$. Suppose $\mathbf{x}^* \in \mathcal{R}^{2L}$, let $\mathbf{x}^*_{\mathbf{B}} = \mathbf{B}^{-1}\mathbf{d}$ and $\mathbf{x}^*_{\mathbf{N}} = 0$. We first prove $\mathbf{x}^*$ is optimal for (10) and hence $\mathbf{x}^*_{\mathcal{I}_1}$ is optimal for (8). Next, we will prove $\mathbf{F}_{\mathcal{B}_1}\mathbf{x}^*_{\mathcal{B}_1} \le \mathbf{b} + \theta\mathbf{1}$, which leads to Proposition 2.

Proof: *Step 1.* To prove $\mathbf{x}^*$ is optimal for equation (10).

It is sufficient to prove $\mathbf{B}^{-1}\mathbf{d} \ge \mathbf{0}$ and $(\mathbf{B}^{-1}\mathbf{N})^T\mathbf{e}_{\mathcal{B}} - \mathbf{e}_{\mathcal{N}} \ge \mathbf{0}$ from Lemma 1, and we will prove $(\mathbf{B}^{-1}\mathbf{N})^T\mathbf{e}_{\mathcal{B}} - \mathbf{e}_{\mathcal{N}} \ge \mathbf{0}$ first. It is easy to show that there always exist permutation matrices $\mathbf{P}_1$ and $\mathbf{P}_2$ to make $\bar{\mathbf{B}} = \mathbf{B}\mathbf{P}_1$ and $\bar{\mathbf{N}} = \mathbf{N}\mathbf{P}_2$, where $\bar{\mathbf{B}}$ and $\bar{\mathbf{N}}$ all have unit-absolute value diagonal elements. Meanwhile, let $\bar{\mathbf{e}}_{\mathbf{B}} = \mathbf{P}_1^T\mathbf{e}_{\mathcal{B}}$ and $\bar{\mathbf{e}}_{\mathbf{N}} = \mathbf{P}_2^T\mathbf{e}_{\mathcal{N}}$. Then $(\mathbf{B}^{-1}\mathbf{N})^T\mathbf{e}_{\mathcal{B}} - \mathbf{e}_{\mathcal{N}} \ge \mathbf{0}$ can be shown equivalent to

$$(\bar{\mathbf{B}}^{-1}\bar{\mathbf{N}})^T\bar{\mathbf{e}}_{\mathcal{B}} - \bar{\mathbf{e}}_{\mathcal{N}} \ge \mathbf{0}. \tag{13}$$

Let $\bar{b}_{ij}$ and $\bar{n}_{ij}$ denote the element of matrices $\bar{\mathbf{B}}$ and $\bar{\mathbf{N}}$ in the ith row and jth column. Then we have

$$\bar{b}_{ij} = \begin{cases} 1 & \text{if } j \in \mathcal{N}_1 \text{ and } i = j \\ 0 & \text{if } j \in \mathcal{N}_1 \text{ and } i \ne j \\ -f_{ij} & \text{if } j \in \mathcal{B}_1 \end{cases}, \tag{14}$$

$$\bar{n}_{ij} = \begin{cases} 1 & \text{if } j \in \mathcal{B}_1 \text{ and } i = j \\ 0 & \text{if } j \in \mathcal{B}_1 \text{ and } i \ne j \\ -f_{ij} & \text{if } j \in \mathcal{N}_1 \end{cases}. \tag{15}$$

Let $\bar{w}_{ij}$ denote the element of matrices $\bar{\mathbf{B}}^{-1}$ in the ith row and jth column. Due to the special structure of $\bar{\mathbf{B}}$, we have

$$\bar{w}_{ij} = \begin{cases} 1 & \text{if } j \in \mathcal{N}_1 \text{ and } i = j \\ 0 & \text{if } j \in \mathcal{N}_1 \text{ and } i \ne j \\ \bar{w}_{ij} & \text{if } j \in \mathcal{B}_1 \end{cases}. \tag{16}$$

Since $\bar{\mathbf{B}}^{-1}\bar{\mathbf{B}} = \mathbf{I}$, combining equations (14) and (16) we have the following equations:

$$\begin{array}{l} (1) \text{ when } i \in \mathcal{N}_1, j \in \mathcal{B}_1 : -f_{ij} - \sum_{k\in\mathcal{B}_1} \bar{w}_{ik}f_{kj} = 0 \\ (2) \text{ when } i \in \mathcal{B}_1, j \in \mathcal{B}_1 \text{ and } i = j : -\sum_{k\in\mathcal{B}_1} \bar{w}_{ik}f_{kj} = 1. \\ (3) \text{ when } i \in \mathcal{B}_1, j \in \mathcal{B}_1 \text{ and } i \ne j : -\sum_{k\in\mathcal{B}_1} \bar{w}_{ik}f_{kj} = 0 \end{array} \tag{17}$$

Let $\bar{\mathbf{D}} = \bar{\mathbf{B}}^{-1}\bar{\mathbf{N}}$. For the special structure of $\bar{\mathbf{e}}_{\mathcal{B}}$ and $\bar{\mathbf{e}}_{\mathcal{N}}$, inequality (13) can be equivalent to:

$$\begin{cases} \mathbf{1}^T_{\mathcal{B}_1}\mathbf{D}_{(\mathcal{B}_1,\mathcal{B}_1)} \le \mathbf{0}^T_{\mathcal{B}_1} \\ \mathbf{1}^T_{\mathcal{B}_1}\mathbf{D}_{(\mathcal{B}_1,\mathcal{N}_1)} \le \mathbf{1}^T_{\mathcal{N}_1} \end{cases}. \tag{18}$$

From equations (14) and (16), we have:

$$\begin{array}{l} \mathbf{D}_{(\mathcal{B}_1,\mathcal{B}_1)} = (\bar{\mathbf{B}}^{-1})_{(\mathcal{B}_1,\mathcal{B}_1)}, \\ \mathbf{D}_{(\mathcal{B}_1,\mathcal{N}_1)} = -(\bar{\mathbf{B}}^{-1})_{(\mathcal{B}_1,\mathcal{B}_1)}\mathbf{F}_{(\mathcal{B}_1,\mathcal{N}_1)}. \end{array} \tag{19}$$

From equation (17), we have:

$$(\bar{\mathbf{B}}^{-1})_{(\mathcal{B}_1,\mathcal{B}_1)}\mathbf{F}_{(\mathcal{B}_1,\mathcal{B}_1)} = -\mathbf{I}. \tag{20}$$

Combining equations (19) and (20), inequality (18) is equivalent to:

$$\begin{cases} \mathbf{1}^T_{\mathcal{B}_1}\mathbf{F}^{-1}_{(\mathcal{B}_1,\mathcal{B}_1)} \ge \mathbf{0}^T_{\mathcal{B}_1} \\ \mathbf{1}^T_{\mathcal{B}_1}\mathbf{F}^{-1}_{(\mathcal{B}_1,\mathcal{B}_1)}\mathbf{F}_{(\mathcal{B}_1,\mathcal{N}_1)} \le \mathbf{1}^T_{\mathcal{N}_1} \end{cases}. \tag{21}$$

Since $\mathbf{F}_{(\mathcal{B}_1,\mathcal{B}_1)}$ is with the same characteristics with $\mathbf{F}$, and the column sums of $\mathbf{F}_{(\mathcal{B}_1,\mathcal{N}_1)}$ are less than 1. From Lemma 3 we can achieve (21), and hence $(\mathbf{B}^{-1}\mathbf{N})^T\mathbf{e}_{\mathcal{B}} - \mathbf{e}_{\mathcal{N}} \ge \mathbf{0}$ is established.

Next we will prove $\mathbf{B}^{-1}\mathbf{d} \ge \mathbf{0}$, which is equivalent to $\bar{\mathbf{B}}^{-1}\mathbf{d} \ge \mathbf{0}$. Recall that $\mathbf{d} = -\mathbf{b} + \theta\mathbf{1} = -\mathbf{F}\mathbf{c}_0 - \mathbf{n} + \theta\mathbf{1}$, and hence $\bar{\mathbf{B}}^{-1}\mathbf{d} \ge \mathbf{0}$ is equivalent to $\bar{\mathbf{B}}^{-1}\mathbf{F}\mathbf{c}_0 \le \bar{\mathbf{B}}^{-1}(\theta\mathbf{1} - \mathbf{n})$.

Let $\mathbf{L} = \bar{\mathbf{B}}^{-1}\mathbf{F}$ and l_{ij} denotes the element of matrices $\mathbf{L}$. From equations (16) and (17), it is easy to verify:

$$l_{ij} = \begin{cases} -1 & \text{if } j \in \mathcal{B}_1 \text{ and } i = j \\ 0 & \text{if } j \in \mathcal{B}_1 \text{ and } i \ne j \\ l_{ij} & \text{if } j \in \mathcal{N}_1 \end{cases}. \tag{22}$$

Recall that the special structure of $\mathbf{c}_0$, we have $\mathbf{E}\mathbf{c}_0 = -\mathbf{c}_0$ and $\bar{\mathbf{B}}^{-1}\mathbf{F}\mathbf{c}_0 \le \bar{\mathbf{B}}^{-1}(\theta\mathbf{1} - \mathbf{n})$ is equivalent to $\bar{\mathbf{B}}^{-1}(\theta\mathbf{1} - \mathbf{n}) \ge -\mathbf{c}_0$, which, considering (16), is also equivalent to:

$$(\bar{\mathbf{B}}^{-1})_{(\mathcal{B}_1,\mathcal{B}_1)}(\theta\mathbf{1} - \mathbf{n})_{\mathcal{B}_1} \ge -(\mathbf{c}_0)_{\mathcal{B}_1} \tag{23}$$

$$(\bar{\mathbf{B}}^{-1})_{(\mathcal{N}_1,:)}(\theta\mathbf{1} - \mathbf{n}) \ge \mathbf{0}. \tag{24}$$

To prove (23), recall that $(\bar{\mathbf{B}}^{-1})_{(\mathcal{B}_1,\mathcal{B}_1)} = -\mathbf{F}^{-1}_{(\mathcal{B}_1,\mathcal{B}_1)}$, and from Lemma 3 it is sufficient to prove $(\theta\mathbf{1} - \mathbf{n})_{\mathcal{B}_1} \le (\mathbf{c}_0)_{\mathcal{B}_1}$

when $\theta \geq \beta\|\mathbf{n}\|_\infty$. The inequality clearly holds because $(\mathbf{c}_0)_{\mathcal{B}_1} \geq (\beta+1)\,\|\mathbf{n}\|_\infty$.

Inequality (24) is equivalent to $(\bar{\mathbf{B}}^{-1})_{(\mathcal{N}_1,\mathcal{B}_1)}(\theta\mathbf{1}-\mathbf{n})_{\mathcal{B}_1} \geq -(\theta\mathbf{1}-\mathbf{n})_{\mathcal{N}_1}$. Combining equations (17); this inequality is also equivalent to $\mathbf{F}_{(\mathcal{N}_1,\mathcal{B}_1)}\mathbf{F}^{-1}_{(\mathcal{B}_1,\mathcal{B}_1)}(\theta\mathbf{1}-\mathbf{n})_{\mathcal{B}_1} \leq (\theta\mathbf{1}-\mathbf{n})_{\mathcal{N}_1}$, which will hold if $\mathbf{F}_{(\mathcal{N}_1,\mathcal{B}_1)}$ $(\theta\mathbf{1}-\mathbf{n})_{\mathcal{B}_1} \leq (\theta\mathbf{1}-\mathbf{n})_{\mathcal{N}_1}$ holds when $\theta \geq \beta\|\mathbf{n}\|_\infty$ from Lemma 3. Since $\|\mathbf{F}_{(\mathcal{N}_1,\mathcal{B}_1)}\|_\infty \leq \frac{1}{2}$, which can be easily verified, we have $\mathbf{F}_{(\mathcal{N}_1,\mathcal{B}_1)}(\theta\mathbf{1}-\mathbf{n})_{\mathcal{B}_1} \leq (\theta\mathbf{1}-\mathbf{n})_{\mathcal{N}_1}$ is established, so is (24). Finally, $\mathbf{B}^{-1}\mathbf{d} \geq \mathbf{0}$ can be reached.

Step 2. To prove $\mathbf{F}_{\mathcal{B}_1}\mathbf{x}^*_{\mathcal{B}_1} \leq \mathbf{b}+\theta\mathbf{1}$.

Recall that $\mathbf{b} = \mathbf{F}\mathbf{c}_0 + \mathbf{n}$ and $\mathbf{d} = -\mathbf{F}\mathbf{c}_0 - \mathbf{n} + \theta\mathbf{1}$, and define $\mathbf{G} = \mathbf{F}_{\mathcal{B}_1}(\mathbf{B}^{-1})_{(\mathcal{B}_1,:)}$, then $\mathbf{F}_{\mathcal{B}_1}\mathbf{x}^*_{\mathcal{B}_1} \leq \mathbf{b}+\theta\mathbf{1}$ is equivalent to $(\mathbf{G}+\mathbf{I})\mathbf{F}\mathbf{c}_0 \geq \mathbf{G}(\theta\mathbf{1}-\mathbf{n}) - (\theta\mathbf{1}-\mathbf{n})$.

Define $\mathbf{H} = \mathbf{G}+\mathbf{I}$, from equations (16) and (17), we have:

$$g_{ij} = \begin{cases} 0 & \text{if } j \in \mathcal{N}_1 \\ -1 & \text{if } j \in \mathcal{B}_1, i \in \mathcal{B}_1 \text{ and } i = j \\ 0 & \text{if } j \in \mathcal{B}_1, i \in \mathcal{B}_1 \text{ and } i \neq j \\ \sum_{k\in\mathcal{B}_1} f_{ik}\bar{w}_{kj} & \text{if } j \in \mathcal{B}_1, i \in \mathcal{N}_1 \end{cases}, \tag{25}$$

$$h_{ij} = \begin{cases} 1 & \text{if } j \in \mathcal{N}_1, \text{ and } i = j \\ 0 & \text{if } j \in \mathcal{N}_1, \text{ and } i \neq j \\ 0 & \text{if } j \in \mathcal{B}_1, i \in \mathcal{B}_1 \\ \sum_{k\in\mathcal{B}_1} f_{ik}\bar{w}_{kj} & \text{if } j \in \mathcal{B}_1, i \in \mathcal{N}_1 \end{cases}. \tag{26}$$

Define $\mathbf{R} = \mathbf{H}\mathbf{F}$, combining equations (17) and (25), we have:

$$r_{ij} = \begin{cases} 0 & \text{if } j \in \mathcal{B}_1 \\ r_{ij} & \text{if } j \in \mathcal{N}_1 \end{cases}. \tag{27}$$

Recall that the special structure of $\mathbf{c}_0$, we have $\mathbf{R}\mathbf{c}_0 = \mathbf{0}$. Therefore, all we need to prove is $\mathbf{G}(\theta\mathbf{1}-\mathbf{n}) - (\theta\mathbf{1}-\mathbf{n}) \leq \mathbf{0}$. Note that:

$$\begin{aligned}\mathbf{G}(\theta\mathbf{1}-\mathbf{n}) &= \mathbf{F}_{\mathcal{B}_1}(\bar{\mathbf{B}}^{-1})_{(\mathcal{B}_1,\mathcal{B}_1)}(\theta\mathbf{1}-\mathbf{n})_{\mathcal{B}_1} \\ &= -\mathbf{F}_{\mathcal{B}_1}\mathbf{F}^{-1}_{(\mathcal{B}_1,\mathcal{B}_1)}(\theta\mathbf{1}-\mathbf{n})_{\mathcal{B}_1} \\ &\leq \mathbf{0}.\end{aligned} \tag{28}$$

Since $\theta \geq \beta\,\|\mathbf{n}\|_\infty$, we have $\mathbf{G}(\theta\mathbf{1}-\mathbf{n}) \leq \mathbf{0} \leq \theta\mathbf{1}-\mathbf{n}$, and hence $\mathbf{F}_{\mathcal{B}_1}\mathbf{x}^*_{\mathcal{B}_1} \leq \mathbf{b}+\theta\mathbf{1}$. □

IV. Proof of the equivalent between equation (5) and (8)

Proof: Proving the equivalence between equations (5) and (8) is actually a problem concerning sensitivity analysis in linear programming, where coefficients in the objective function are perturbed. Under the assumptions in Proposition 2, we know, from Appendix A.III, that the basic variables set $\mathcal{B} = \bar{\mathcal{Z}} \cup (\mathcal{Z}+L)$ and nonbasic variables set $\mathcal{N} = \mathcal{Z} \cup (\bar{\mathcal{Z}}+L)$ are optimal for problem (5). Suppose that x^* is the optimal solution of problem (5), then we have $x^*_{\mathcal{B}} = \mathbf{B}^{-1}\mathbf{d}$ and $x^*_{\mathcal{N}} = 0$. Now we prove x^* is also optimal to problem (8).

Problem (8) can also be transformed into equations (10)–(12), with $\mathbf{e} = [-\mathbf{1}^T\mathbf{F}, \mathbf{0}^T]^T$. Since $x^*_{\mathcal{B}} = \mathbf{B}^{-1}\mathbf{d} \geq \mathbf{0}$ and $x^*_{\mathcal{N}} = 0$, it is obvious that x^* is feasible for problem (8). If we can further prove that $(\mathbf{B}^{-1}\mathbf{N})^T\mathbf{e}_{\mathcal{B}} - \mathbf{e}_{\mathcal{N}} \geq \mathbf{0}$, the conclusion that x^* is also optimal to problem (8) is established.

Similar to (21), to prove $(\mathbf{B}^{-1}\mathbf{N})^T\mathbf{e}_{\mathcal{B}} - \mathbf{e}_{\mathcal{N}} \geq \mathbf{0}$ is equivalent to prove

$$\begin{cases} \mathbf{1}^T\mathbf{F}_{\mathcal{B}_1}\mathbf{F}^{-1}_{(\mathcal{B}_1,\mathcal{B}_1)} \geq \mathbf{0}^T_{\mathcal{B}_1} \\ \mathbf{1}^T\mathbf{F}_{\mathcal{B}_1}\mathbf{F}^{-1}_{(\mathcal{B}_1,\mathcal{B}_1)}\mathbf{F}_{(\mathcal{B}_1,\mathcal{N}_1)} \leq \mathbf{1}^T\mathbf{F}_{\mathcal{N}_1} \end{cases}. \tag{29}$$

For the first inequality in equation (29), note that

$$\begin{aligned} &\mathbf{1}^T\mathbf{F}_{\mathcal{B}_1}\mathbf{F}^{-1}_{(\mathcal{B}_1,\mathcal{B}_1)} \\ &= \mathbf{1}^T[\mathbf{F}^T_{(\mathcal{B}_1,\mathcal{B}_1)}\ \mathbf{F}^T_{(\mathcal{N}_1,\mathcal{B}_1)}]^T\mathbf{F}^{-1}_{(\mathcal{B}_1,\mathcal{B}_1)} \\ &= \mathbf{1}^T[\mathbf{I}\ \mathbf{F}^T_{(\mathcal{N}_1,\mathcal{B}_1)}]^T\mathbf{F}^{-1}_{(\mathcal{B}_1,\mathcal{B}_1)}. \end{aligned} \tag{30}$$

It is easy to show $\mathbf{F}_{(\mathcal{N}_1,\mathcal{B}_1)}\mathbf{F}^{-1}_{(\mathcal{B}_1,\mathcal{B}_1)} \geq \mathbf{0}$. Suppose to the contrary that $\mathbf{F}_{(\mathcal{N}_1,\mathcal{B}_1)}\mathbf{F}^{-1}_{(\mathcal{B}_1,\mathcal{B}_1)} < \mathbf{0}$, after multiplying $\mathbf{F}_{(\mathcal{B}_1,\mathcal{B}_1)}$ to the right side we achieve $\mathbf{F}_{(\mathcal{N}_1,\mathcal{B}_1)}\mathbf{F}^{-1}_{(\mathcal{B}_1,\mathcal{B}_1)}\mathbf{F}_{(\mathcal{B}_1,\mathcal{B}_1)} < \mathbf{0}\ \mathbf{F}_{(\mathcal{B}_1,\mathcal{B}_1)}$, i.e. $\mathbf{F}_{(\mathcal{N}_1,\mathcal{B}_1)} < 0$, which is impossible. Now we can conclude (30) is positive.

For the second inequality in equation (29), it is established if we can prove:

$$\begin{cases} \mathbf{1}^T_{\mathcal{B}_1}\mathbf{F}_{(\mathcal{B}_1,\mathcal{B}_1)}\mathbf{F}^{-1}_{(\mathcal{B}_1,\mathcal{B}_1)}\mathbf{F}_{(\mathcal{B}_1,\mathcal{N}_1)} \leq \mathbf{1}^T_{\mathcal{B}_1}\mathbf{F}_{(\mathcal{B}_1,\mathcal{N}_1)} \\ \mathbf{1}^T_{\mathcal{N}_1}\mathbf{F}_{(\mathcal{N}_1,\mathcal{B}_1)}\mathbf{F}^{-1}_{(\mathcal{B}_1,\mathcal{B}_1)}\mathbf{F}_{(\mathcal{B}_1,\mathcal{N}_1)} \leq \mathbf{1}^T_{\mathcal{N}_1}\mathbf{F}_{(\mathcal{N}_1,\mathcal{N}_1)} \end{cases}. \tag{31}$$

The first inequality in equation (31) holds clearly. For the second inequality, since $\mathbf{F}^{-1}_{(\mathcal{B}_1,\mathcal{B}_1)}\mathbf{F}_{(\mathcal{B}_1,\mathcal{N}_1)} \geq \mathbf{0}$, $\mathbf{1}^T_{\mathcal{N}_1}\mathbf{F}_{(\mathcal{N}_1,\mathcal{B}_1)} \leq \mathbf{1}^T_{\mathcal{B}_1}$ and $\mathbf{1}^T_{\mathcal{B}_1}\mathbf{F}^{-1}_{(\mathcal{B}_1,\mathcal{B}_1)} \leq \mathbf{1}^T_{\mathcal{B}_1}$, we have $\mathbf{1}^T_{\mathcal{N}_1}\mathbf{F}_{(\mathcal{N}_1,\mathcal{B}_1)}\mathbf{F}^{-1}_{(\mathcal{B}_1,\mathcal{B}_1)}\mathbf{F}_{(\mathcal{B}_1,\mathcal{N}_1)} \leq \mathbf{1}^T_{\mathcal{B}_1}\mathbf{F}^{-1}_{(\mathcal{B}_1,\mathcal{B}_1)}\mathbf{F}_{(\mathcal{B}_1,\mathcal{N}_1)} \leq \mathbf{1}^T_{\mathcal{B}_1}\mathbf{F}_{(\mathcal{B}_1,\mathcal{N}_1)} < \mathbf{1}^T_{\mathcal{N}_1}\mathbf{F}_{(\mathcal{N}_1,\mathcal{N}_1)}$, and hence equation (31) holds and so does (29). Therefore $(\mathbf{B}^{-1}\mathbf{N})^T\mathbf{e}_{\mathcal{B}} - \mathbf{e}_{\mathcal{N}} \geq \mathbf{0}$ is established and x^* is also optimal for problem (8). □

Appendix B

First we show two obvious but useful lemmas, both of which are easy to verify.

Lemma 7: *If $\mathbf{A} \in \mathcal{R}^{n\times n}$ is nonsingular with all its row's sums equal to s, then the row's sums of $\mathbf{A}^{-1}$ all equal to $\frac{1}{s}$.*

Lemma 8: *Suppose $\mathbf{A}_0, \triangle\mathbf{A}, \mathbf{B}_0, \triangle\mathbf{B} \in \mathcal{R}^{n\times n}$, and both $\mathbf{A}_0$ and $\mathbf{A}_0+\triangle\mathbf{A}$ are nonsingular, besides, suppose $\mathbf{A}_0^{-1} = \mathbf{B}_0$, then we have $(\mathbf{A}_0+\triangle\mathbf{A})^{-1} = \mathbf{B}_0+\triangle\mathbf{B}$, where $\triangle\mathbf{B} = -\mathbf{B}_0\triangle\mathbf{A}\mathbf{B}_0 - \mathbf{B}_0\triangle\mathbf{A}\triangle\mathbf{B}$.*

The following lemma is important to later theoretical analysis.

Lemma 9: *Suppose $\mathbf{F}$ is nonsingular. Split $\mathbf{F}$ into $\mathbf{F} = \mathbf{F}_0 + \triangle\mathbf{F}$, where $\mathbf{F}_0$ is nonsingular with all row's sums equal. If $\|\triangle\mathbf{F}\|_\infty \leq \alpha\|\mathbf{F}_0\|_\infty$, where $\alpha \in [0, \frac{1}{2})$, then when $\theta \geq \frac{1}{1-2\alpha}\|\mathbf{n}\|_\infty$, we have $\mathbf{F}^{-1}(\theta\mathbf{1}-\mathbf{n}) \geq \mathbf{0}$.*

Proof: Suppose $\mathbf{G}_0, \triangle\mathbf{G} \in \mathcal{R}^{L\times L}$, and $\mathbf{G}_0 = \mathbf{F}_0^{-1}$, $(\mathbf{G}_0+\triangle\mathbf{G}) = (\mathbf{F}_0+\triangle\mathbf{F})^{-1}$. By Lemma 5, we have $\triangle\mathbf{G} = -\mathbf{G}_0\triangle\mathbf{F}\mathbf{G}_0 - \mathbf{G}_0\triangle\mathbf{F}\triangle\mathbf{G}$. It is easy to achieve

$$(1-\|\mathbf{G}_0\|_\infty\|\triangle\mathbf{F}\|_\infty)\|\triangle\mathbf{G}\|_\infty \leq \|\mathbf{G}_0\|_\infty\|\triangle\mathbf{F}\|_\infty\|\mathbf{G}_0\|_\infty.$$

Note that $\|\mathbf{F}_0\|_\infty \|\mathbf{G}_0\|_\infty = 1$ and $1 - \|\mathbf{G}_0\|_\infty \|\triangle\mathbf{F}\|_\infty > 0$, and hence we can achieve $\|\triangle\mathbf{G}\|_\infty \leq \frac{\alpha}{1-\alpha}\|\mathbf{G}_0\|_\infty < \|\mathbf{G}_0\|_\infty$ for $\alpha \in [0, \frac{1}{2})$.

On the other hand, note that $\mathbf{F}^{-1}(\theta\mathbf{1} - \mathbf{n}) = (\mathbf{G}_0 + \triangle\mathbf{G})(\theta\mathbf{1} - \mathbf{n})$, and to prove $\mathbf{F}^{-1}(\theta\mathbf{1} - \mathbf{n}) \geq \mathbf{0}$ it is sufficient to prove $(\|\mathbf{G}_0\|_\infty + \|\triangle\mathbf{G}\|_\infty)\|\mathbf{n}\|_\infty \leq (\|\mathbf{G}_0\|_\infty - \|\triangle\mathbf{G}\|_\infty)\ \theta$, which is equivalent to

$$\theta \geq (1 + \|\mathbf{F}_0\|_\infty \|\triangle\mathbf{G}\|_\infty)\|\mathbf{n}\|_\infty / (1 - \|\mathbf{F}_0\|_\infty \|\triangle\mathbf{G}\|_\infty).$$

It is easy to show $(1 + \|\mathbf{F}_0\|_\infty \|\triangle\mathbf{G}\|_\infty)\|\mathbf{n}\|_\infty / (1 - \|\mathbf{F}_0\|_\infty \|\triangle\mathbf{G}\|_\infty) \leq \|\mathbf{n}\|_\infty / (1 - 2\alpha)$, and (32) holds for $\theta \geq \frac{1}{(1-2\alpha)}\|\mathbf{n}\|_\infty$. □

Now we prove Lemma 6.

Proof: Split $\mathbf{F}$ into $\mathbf{F} = \mathbf{F}_0 + \triangle\mathbf{F}$, where in $\mathbf{F}_0$ all row sums equal $s_{\max}$ and all diagonal elements equal one, and $\triangle\mathbf{F}$ is nonnegative with zero diagonal elements. It is noted that such split will always exists. Let $s_{\max} = 1 + x$ and $s_{\min} = 1 + y$, we have $1 > x > y \geq 0$. Note that $\|\triangle\mathbf{F}\|_\infty = (1 + x) - (1 + y) = x - y$, thus we have $\frac{1}{2}\|\mathbf{F}_0\|_\infty - \|\triangle\mathbf{F}\|_\infty = \frac{1}{2}s_{\max} - (x - y) = \frac{1}{2}(1 - x) + y > 0$, i.e., $\|\triangle\mathbf{F}\|_\infty < \frac{1}{2}\|\mathbf{F}_0\|_\infty$. That is to say, the perturbation matrix $\triangle\mathbf{F}$ satisfies the assumptions of Lemma 6. Let $\mathbf{n} = \mathbf{0}$, we have $\mathbf{F}^{-1}\mathbf{1} \geq \mathbf{0}$. Because if $\mathbf{F}\mathbf{x} = \mathbf{1}$ and $\mathbf{x} \geq \mathbf{0}$, then $\mathbf{x} \leq \mathbf{1}$, and hence $\mathbf{F}^{-1}\mathbf{1} \leq \mathbf{1}$.

Second, since $\frac{y}{x} > \frac{1}{\beta}$, $\alpha = \|\triangle\mathbf{F}\|_\infty / \|\mathbf{F}_0\|_\infty = (x - y)/(1 + x) < (1 - \frac{1}{\beta}\ )\frac{x}{1+x} < \frac{1}{2}(1 - \frac{1}{\beta}\ )$. From Lemma 6 WE know if $\theta \geq \beta\|\mathbf{n}\|_\infty$, we have $\mathbf{F}^{-1}(\theta\mathbf{1} - \mathbf{n}) \geq \mathbf{0}$. And $\mathbf{F}^{-1}(\theta\mathbf{1} - \mathbf{n}) \leq \mathbf{1}$ is easy to verify. □