Decay of Correlation in Network Flow Problems

Patrick Rebeschini
Yale Institute for Network Science
17 Hillhouse Avenue, Room 324
New Haven, CT 06511
Email: patrick.rebeschini@yale.edu

Sekhar Tatikonda
Yale Institute for Network Science
17 Hillhouse Avenue, Room 338
New Haven, CT 06511
Email: sekhar.tatikonda@yale.edu

***Abstract*—We develop a general theory for the local sensitivity of optimal points of constrained network optimization problems under perturbations of the constraints. For the network flow problem, we show that local perturbations on the constraints have an impact on the components of the optimal point that decreases exponentially with the graph-theoretical distance. The exponential rate is controlled by the spectral radius of a substochastic transition matrix of a killed random walk associated to the network. For graphs where this spectral radius is well-behaved (bounded, for instance) as a function of the dimension of the network, our theory yields the first-known incarnation of the decay of correlation principle in constrained optimization.**

***Index Terms*—sensitivity of optimal point, decay of correlation, network flow, killed random walk**

I. Introduction

Consider the problem of minimizing a twice continuously differentiable strongly convex function $f : x \to f(x)$ given the equality constraint $Ax = b$, where A is a full row rank matrix. Let $x^\star(b)$ denote the unique solution of this problem, as a function of the constraint vector b. How does $x^\star(b)_i$ — the i-th component of $x^\star(b)$ — behave upon perturbation of b_a — the a-th component of b? Results on the sensitivity analysis for optimization procedures are typically stated only with respect to the optimal objective function, i.e., $f(x^\star(b))$, not with respect to the point where the optimum is attained, i.e., $x^\star(b)$. But can we express $\frac{\partial x^\star(b)_i}{\partial b_a}$ itself as a function of f, A, and b, and how does this quantity behave with respect to i and a, and the network topology? These are the type of questions that we address, from the point of view of network locality and decay of correlation. Our main contributions are:

1) (General theory) Provide the first systematic analysis of decay of correlation for optimization procedures.
2) (Network flow problem) Show that the optimal network flow problem *structurally* exhibits decay of correlation. Provide a characterization of correlations in terms of spectral analysis of killed random walks.

We consider the widely-studied class of optimal network flow problems that has been fundamental in the development of the theory of polynomial-times algorithms for optimizations (see [1] and references therein, or [2] for book reference). Here a directed graph $G = (V, E)$ is given with its structure encoded in the vertex-edge incidence matrix A. To each vertex $a \in V$ is associated an external flow $b_a \in \mathbb{R}$. To each edge $i \in E$ is associated a cost function $f_i : x_i \in \mathbb{R} \to f_i(x_i) \in \mathbb{R}$, where x_i is the flow along edge i. Here $x^\star(b) = (x^\star(b)_i)_{i\in E}$ represents the flow that minimizes the total cost $\sum_{i\in E} f_i$ in the network and that satisfies the conservation law $Ax = b$ so that at each vertex the total flow is zero, where $b = (b_a)_{a\in V}$ is the external flow. In this context, the quantity $\frac{\partial x^\star(b)_i}{\partial b_a}$ can be interpreted as a measure of the "correlation" between edge i and vertex a. One of our main results shows that the magnitude of the correlation $|\frac{\partial x^\star(b)_i}{\partial b_a}|$ is upper bounded by a quantity that decays exponentially with the graph-theoretical distance between i and a, with rate given by the spectral radius of a substochastic transition matrix of a killed random walk associated to the network. For graphs where this spectral radius is well-behaved (bounded, for instance) as a function of the dimension of the network, our theory yields the first-known incarnation of the decay of correlation principle in constrained optimization.

The concept of decay of correlation has been widely studied in statistical mechanics and probability theory, starting with the seminal work of Dobrushin [3] to investigate the problem of uniqueness of Gibbs measures on infinite graphs (for book references see [4] and [5]). In this setting, decay of correlation characterizes the effective neighborhood dependency of random variables in a probabilistic network. Since the work of Dobrushin, this concept has found many applications beyond statistical physics, see [6] for instance. However, even when the underlying problem is completely deterministic, such as in classical optimization procedures, the decay of correlation property is typically established upon endowing the model with a probabilistic structure: randomness and independence are embedded in various ways so that the desired decay of correlation property can be established and exploited. The only case we are aware of where the decay of correlation property has been explicitly considered in a purely deterministic setting in optimization is treated in [7]. In this paper the authors use decay of correlation to prove the convergence of the min-sum message passing algorithm to solve the class of separable *unconstrained* convex optimization problems. Yet, decay of correlation is simply regarded as a tool to prove convergence guarantees for the specific algorithm at hand, and no general theory is built around it. On the other hand, the need to address diverse large-scale graphical models applications in the optimization and machine learning domain prompts to investigate the *foundations* of deterministic correlation decay, and to develop a general theory that can then inspire a principled use of this concept for local distributed algorithms.

Our results represent a first step in this direction. The general characterization that we give to $\frac{\partial x^\star(b)_i}{\partial b_a}$ (Theorem 2.1 below) can be interpreted as a first instance of comparison theorems for constrained optimization procedures, along the lines of the comparison theorems established in probability theory to capture stochastic decay of correlation and control the difference of high-dimensional distributions (see [8] and [3]).

The rest of the paper is organized as follows. In Section II we develop the local sensitivity analysis to characterize the correlation term $\frac{\partial x^\star(b)_i}{\partial b_a}$ as a function of f, A, and b for generic equality-constrained convex problems, making explicit the interplay between the Hessian of f and the structure of the constraint matrix A. In Section III we introduce the minimum-cost network flow problem, and in Section IV we specialize the local sensitivity analysis of Section II to this problem. We show that the optimal network flow problem *structurally* exhibits decay of correlation (point-to-point, point-to-set, and set-to-point) that can be captured via the spectral analysis of killed random walks on graphs.

II. Local sensitivity for constrained optimization

Let $\mathcal{V}$ be a finite set — to be interpreted as the "variable set" — with cardinality $|\mathcal{V}|$, and let $f : \mathbb{R}^\mathcal{V} \to \mathbb{R}$ be a strongly convex function, twice continuously differentiable. Let $\mathcal{F}$ be a finite set — to be interpreted as the "factor set" — with cardinality $|\mathcal{F}|$, and let $A \in \mathbb{R}^{\mathcal{F}\times\mathcal{V}}$. Consider the following optimization problem over $x \in \mathbb{R}^\mathcal{V}$:

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & Ax = b, \end{array}$$

where $b \in \mathbb{R}^\mathcal{F}$. Assume that $|\mathcal{F}| \le |\mathcal{V}|$ and that A has full rank $|\mathcal{F}|$, so that the equality constraint represents an independent set of equations. Let the function f and the matrix A be fixed, and let us consider the solution of the optimization problem above as a function of the vector b. It is easy to verify that strong convexity implies that this problem has a unique optimal solution. For each $b \in \mathbb{R}^\mathcal{F}$, let

$$x^\star(b) := \arg\min \left\{ f(x) : x \in \mathbb{R}^\mathcal{V}, Ax = b \right\}.$$

The following theorem establishes that the function $x^\star$ is continuously differentiable, and it provides a local characterization of the way a perturbation of the constraint vector b affects the optimal solution $x^\star(b)$. In textbooks, results on the sensitivity analysis for optimization procedures as a function of the parameters of the model are typically stated only with respect to the optimal objective function, i.e., $f(x^\star(b))$, not with respect to the point where the optimum is attained, i.e., $x^\star(b)$, as in the following theorem.

Theorem 2.1: Let $f : \mathbb{R}^\mathcal{V} \to \mathbb{R}$ be a strongly convex function, twice continuously differentiable. Let $A \in \mathbb{R}^{\mathcal{F}\times\mathcal{V}}$ be a full rank matrix, with $|\mathcal{F}| \le |\mathcal{V}|$. For each $b \in \mathbb{R}^\mathcal{F}$, let $H(b) := \nabla^2 f(x^\star(b))$, $L(b) := AH(b)^{-1}A^T$, where A^T is the transpose of A, and define

$$D(b) := H(b)^{-1}A^T L(b)^{-1}.$$

Then, $x^\star$ is continuously differentiable, and for each $i \in \mathcal{V}$, $a \in \mathcal{F}$, and $b \in \mathbb{R}^\mathcal{F}$, we have

$$\frac{\partial x^\star(b)_i}{\partial b_a} = D(b)_{ia}.$$

Proof Define the function Φ from $\mathbb{R}^\mathcal{V} \times \mathbb{R}^\mathcal{F}$ to $\mathbb{R}^\mathcal{V} \times \mathbb{R}^\mathcal{F}$ as

$$\Phi \begin{pmatrix} x \\ \nu \end{pmatrix} := \begin{pmatrix} \nabla f(x) + A^T \nu \\ Ax \end{pmatrix}.$$

For each $b \in \mathbb{R}^\mathcal{F}$, the minimizer $x^\star(b)$ satisfies the Karush-Kuhn-Tucker (KKT) conditions

$$\Phi \begin{pmatrix} x^\star(b) \\ \nu^\star(b) \end{pmatrix} = \begin{pmatrix} 0 \\ b \end{pmatrix}, \tag{1}$$

where $\nu^\star(b) \in \mathbb{R}^\mathcal{F}$ is the optimal dual variable. By Hadamard global inverse theorem (see [9] for instance), it can be shown that the function Φ is a C^1 diffeomorphism, namely, it is continuously differentiable, bijective, and its inverse is also continuously differentiable. In particular, this means that the functions $x^\star : b \in \mathbb{R}^\mathcal{F} \to x^\star(b)$ and $\nu^\star : b \in \mathbb{R}^\mathcal{F} \to \nu^\star(b)$ are continuously differentiable. Differentiating both sides of (1) with respect to b_a, $a \in \mathcal{F}$, by the chain rule we find

$$J(b) \begin{pmatrix} \frac{\partial x^\star(b)}{\partial b_a} \\ \frac{\partial \nu^\star(b)}{\partial b_a} \end{pmatrix} = \begin{pmatrix} 0 \\ e_a \end{pmatrix}, \qquad J(b) := \begin{pmatrix} H(b) & A^T \\ A & 0 \end{pmatrix},$$

where $H(b) := \nabla^2 f(x^\star(b))$ and $(e_a)_{a'} := \mathbf{1}_{a=a'}$, with $\mathbf{1}$ being the indicator function. As the function f is strongly convex, the Hessian $\nabla^2 f(x)$ is positive definite for every $x \in \mathbb{R}^\mathcal{V}$, hence it is invertible. As A is full rank, the quantity $L(b) := AH(b)^{-1}A^T \in \mathbb{R}^{\mathcal{F}\times\mathcal{F}}$ is positive definite for every $b \in \mathbb{R}^\mathcal{F}$. To see this, let $y \in \mathbb{R}^\mathcal{F}, y \ne 0$. Since A^T has full column rank, we have $z = A^T y \ne 0$, and as $H(b)$ is positive definite we have $y^T AH(b)^{-1}A^T y = z^T H(b)^{-1} z > 0$. Therefore, $L(b)$ is invertible, and the inverse of the block matrix $J(b)$ reads (not writing the dependence on b)

$$J^{-1} = \begin{pmatrix} H^{-1} - H^{-1}A^T L^{-1} A H^{-1} & H^{-1}A^T L^{-1} \\ L^{-1}AH^{-1} & -L^{-1} \end{pmatrix}.$$

As

$$\begin{pmatrix} \frac{\partial x^\star(b)}{\partial b_a} \\ \frac{\partial \nu^\star(b)}{\partial b_a} \end{pmatrix} = J(b)^{-1} \begin{pmatrix} 0 \\ e_a \end{pmatrix},$$

we have $\frac{\partial x^\star(b)}{\partial b_a} = H(b)^{-1}A^T L(b)^{-1} e_a$ and $\frac{\partial x^\star(b)_i}{\partial b_a} = \left(H(b)^{-1}A^T L(b)^{-1}\right)_{ia} = D(b)_{ia}$ for $i \in \mathcal{V}$.

The quantity $D(b)_{ia}$ in Theorem 2.1 measures the impact that a perturbation of the a-th component of the constraint vector b has to the i-th component of the solution $x^\star(b)$. Thus, $D(b)_{ia}$ describes the "correlation" between i and a.

III. Optimal Network Flow

We now introduce the minimum-cost network flow problem, a cornerstone in the development of the theory of polynomial-times algorithms for optimizations. We refer to [1] for an account of the importance that this problem has had in the field of optimization, and to [2] for book reference. The problem formulation as here presented is taken from Chapter 10 in [10].

Consider a connected[1] directed graph $\bar{G} := (\bar{V}, E)$, with no self-edges and no multiple edges, with vertex set $\bar{V}$ and edge set E. Assume $|\bar{V}| - 1 \le |E|$. For each $i \in E$ let x_i denote the flow on edge i, with $x_i > 0$ if the flow is in the direction of the edge, $x_i < 0$ if the flow is in the direction opposite the edge. For each $a \in \bar{V}$ let $\bar{b}_a$ be a given external flow (source or sink) on the vertex a, with $\bar{b}_a > 0$ if the flow enters the vertex, $\bar{b}_a < 0$ if the flow leaves the vertex. Assume that the total of the source flows equals the total of the sink flows, that is, $\sum_{a \in \bar{V}} \bar{b}_a = 0$. We assume that the flow satisfies a conservation equation so that at each vertex the total flow is zero. This conservation law can be expressed as $\bar{A}x = \bar{b}$, where $\bar{A} \in \mathbb{R}^{\bar{V} \times E}$ is the *vertex-edge incidence matrix* defined for each $a \in \bar{V}$ and $i \in E$ as

$$\bar{A}_{ai} := \begin{cases} 1 & \text{if edge } i \text{ leaves node } a, \\ -1 & \text{if edge } i \text{ enters node } a, \\ 0 & \text{otherwise,} \end{cases}$$

and $\bar{b} \in \mathbb{R}^{\bar{V}}$ represents the external flow. The conservation equations represented by $\bar{A}x = \bar{b}$ are redundant as we clearly have $1^T \bar{A} = 0^T$ (and $1^T \bar{b} = 0$), where 1 and 0 are the all-ones and all-zeros vectors, respectively. To obtain an independent set of equations we can disregard any of them. To this end, henceforth, fix $\bar{a} \in \bar{V}$, let $V := \bar{V} \setminus \bar{a}$, and define the pair $G := (V, E)$. Note that G no longer has a graph structure as we have removed $\bar{a}$ and now there are some edges in E that do not connect pairs of elements in V. Let $A \in \mathbb{R}^{V \times E}$ be the restriction of $\bar{A}$ on G, that is, the matrix $\bar{A}$ with the $\bar{a}$-th row removed, and define $\bar{b}_V \in \mathbb{R}^V$ analogously, removing from $\bar{b}$ the entry associated to $\bar{a}$. The flow conservation is now equivalently written as $Ax = \bar{b}_V$. Clearly, the matrix A has full row rank $|V| = |\bar{V}| - 1$. Consider the following problem.

Optimal network flow problem For each edge $i \in E$ let $f_i : \mathbb{R} \to \mathbb{R}$ be its associated cost function, assumed to be strongly convex and twice continuously differentiable. Let $A \in \mathbb{R}^{V \times E}$ be defined as above, and let $b = (b_a)_{a \in V} \in \mathbb{R}^V$ be given. Then the *optimal network flow problem* is

$$\begin{aligned} \text{minimize} \quad & f(x) := \sum_{i \in E} f_i(x_i) \\ \text{subject to} \quad & Ax = b. \end{aligned} \tag{2}$$

Notice that the structure $G = (V, E)$ behind the constraint equations $Ax = b$ can be interpreted as a factor graph with variable set $\mathcal{V} := E$, factor set $\mathcal{F} := V$, and where there is an (undirected) edge between $i \in \mathcal{V}$ and $a \in \mathcal{F}$ if and only if $A_{ai} \neq 0$. This interpretation immediately yields the following convenient notion of neighborhoods. For $i \in E$, $a \in V$, let

$$\partial i := \{a \in V : A_{ai} \neq 0\}, \qquad \partial a := \{i \in E : A_{ai} \neq 0\}.$$

The original graph $\bar{G} = (\bar{V}, E)$ can also be interpreted as a factor graph, with variable set $\mathcal{V}$, factor set $\bar{\mathcal{F}} := \mathcal{F} \cup \bar{a} \equiv \bar{V}$, and where there is an (undirected) edge between $i \in \mathcal{V}$ and $a \in \bar{\mathcal{F}}$ if and only if $\bar{A}_{ai} \neq 0$. For each $i \in E$ and $a \in \bar{V}$, let

$$\bar{\partial} i := \{a \in V : \bar{A}_{ai} \neq 0\}, \qquad \bar{\partial} a := \{i \in E : \bar{A}_{ai} \neq 0\}.$$

As the objective function f in (2) is separable, without loss of generality we assume that the subgraph $(V, E \setminus \bar{\partial}\bar{a})$ is connected. Otherwise, we can break the optimization problem (2) into its disconnected parts, and treat each of them separately.

IV. Local sensitivity: killed random walks

In this section we use the local sensitivity analysis provided by Theorem 2.1 to investigate the quantity $\frac{\partial x^\star(b)_i}{\partial b_a}$, $i \in E$, $a \in V$, in the optimal network flow problem. For simplicity of notation, henceforth we neglect to write explicitly the dependence on b. For instance, we write $x^\star$ to indicate $x^\star(b)$. We introduce a few quantities of interest. For each $i \in E$, let

$$w_i := \left(\frac{\partial^2 f_i(x_i^\star)}{\partial x_i^2} \right)^{-1} > 0,$$

which is strictly positive as f_i is strongly convex by assumption. Let $W \in \mathbb{R}^{V \times V}$ be the symmetric matrix defined as follows, for each $a, a' \in V$,

$$W_{aa'} := \begin{cases} w_i & \text{if } i = (a, a') \in E \text{ or } i = (a', a) \in E, \\ 0 & \text{otherwise,} \end{cases}$$

and let $D \in \mathbb{R}^{V \times V}$ be the diagonal matrix with entries, for each $a \in V$, $d_a := D_{aa} := \sum_{i \in \partial a} w_i$. Define $L := D - W$, or, entry-wise for each $a, a' \in V$,

$$L_{aa'} = \begin{cases} \sum_{i \in \partial a} w_i & \text{if } a = a', \\ -w_i & \text{if } i = (a, a') \in E \text{ or } i = (a', a) \in E, \\ 0 & \text{otherwise.} \end{cases}$$

From Theorem 2.1 we immediately have the following result on the sensitivity of the optimal flow problem.

Lemma 4.1: For each $a \in V$, $i = (a', a'') \in E$ we have

$$\frac{\partial x_i^\star}{\partial b_a} = w_i \left\{ L^{-1}_{a'a} - L^{-1}_{a''a} \right\},$$

where we adopt the convention that $L^{-1}_{\bar{a}a} := 0$ for any $a \in V$.

We now show how we can write $\frac{\partial x_i^\star}{\partial b_a}$ in terms of Neumann power series of sub-stochastic matrices that have an interpretation in terms of killed random walks on the undirected graph $(V, |E \setminus \bar{\partial}\bar{a}|)$, as we discuss next. Here the notation $|E \setminus \bar{\partial}\bar{a}|$ is used to denote the undirected edge set formed by the directed edge set $E \setminus \bar{\partial}\bar{a}$ by mapping each directed edge to an undirected edge that connects the same pair of nodes.

Let us first recall a few facts on sub-stochastic matrices.

Sub-stochastic matrix A *sub-stochastic matrix* is a matrix with non-negative entries that has row sums less than or equal to 1, with at least one row sum less than 1.

Given a sub-stochastic matrix $P \in \mathbb{R}^{V \times V}$, define the distance between $a \in V$ and $a' \in V$ as follows

$$d(a, a') := \inf\{t \ge 0 : P^t_{aa'} \neq 0\}.$$

[1] A directed graph is connected if the *undirected* graph naturally associated to it is connected.

The matrix P is *irreducible* (in the sense of Markov chains) if $d(a, a') < \infty$ for each $a, a' \in V$.

Lemma 4.2: Let P be a sub-stochastic matrix, and let ρ be its spectral radius. Then, $\rho \leq 1$. If P is irreducible then $\rho < 1$.

Proof See Corollary 6.2.28 in [11], for instance.

Henceforth, let $P := D^{-1}W \in \mathbb{R}^{V\times V}$, i.e., for $a, a' \in V$,

$$P_{aa'} = \begin{cases} \frac{w_i}{\sum_{j\in\partial a} w_j} & \text{if } i = (a, a') \in E \text{ or } i = (a', a) \in E, \\ 0 & \text{otherwise.} \end{cases}$$

Lemma 4.3: The matrix P is sub-stochastic and for each $a \in V$ and $i = (a', a'') \in E$ we have

$$\frac{\partial x_i^\star}{\partial b_a} = \frac{w_i}{d_a} \sum_{t=0}^{\infty} \left\{ (P^t)_{a'a} - (P^t)_{a''a} \right\},$$

where we adopt the convention that $P_{\bar{a}a} := 0$ for any $a \in V$.

Proof The matrix P is sub-stochastic as, clearly, if $a \notin \partial\bar{a}$ then $\sum_{a'\in V} P_{aa'} = 1$, while if $a \in \partial\bar{a}$ then $\sum_{a'\in V} P_{aa'} < 1$. As $(V, E \setminus \partial\bar{a})$ is connected by assumption, then P is irreducible and by Lemma 4.2 the spectral radius of P is strictly less than 1, so that the Neumann series $\sum_{t=0}^{\infty} P^t$ converges. Then, the Neumann series expansion for L^{-1} reads

$$L^{-1} = \sum_{t=0}^{\infty} (I - D^{-1}L)^t D^{-1} = \sum_{t=0}^{\infty} P^t D^{-1}.$$

The final statement follows immediately from Theorem 4.1.

The matrix P can be interpreted as the transition matrix of the killed random walk on the undirected weighted graph $(V, |E \setminus \partial\bar{a}|, W)$ that is obtained by creating a cemetery at $\bar{a}$ in the (regular) random walk on the weighted graph $(V \cup \bar{a}, |E|, \bar{W})$, where the weight matrix $\bar{W} \in \mathbb{R}^{(V\cup\bar{a})\times(V\cup\bar{a})}$ is defined, for each $a, a' \in (V \cup \bar{a})$, as

$$\bar{W}_{aa'} := \begin{cases} w_i & \text{if } i = (a, a') \in E \text{ or } i = (a', a) \in E, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

The (stochastic) transition matrix of the random walk on $(V \cup \bar{a}, |E|, \bar{W})$ is given by the standard diffusion operator $\bar{P} := \bar{D}^{-1}\bar{W}$, where $\bar{D} \in \mathbb{R}^{(V\cup\bar{a})\times(V\cup\bar{a})}$ is a diagonal matrix with entries given, for each $a \in (V \cup \bar{a})$, by $d_a := \bar{D}_{aa} := \sum_{i\in\bar{\partial}a} w_i$. Creating a cemetery at $\bar{a}$ means modifying the walk so that $\bar{a}$ becomes a recurrent state, i.e., once the walk is in state $\bar{a}$ it will go back to $\bar{a}$ with probably 1. This is clearly done by replacing the $\bar{a}$-th row of $\bar{P}$ by a row with zeros everywhere but in the $\bar{a}$-th coordinate, where the entry is equal to 1. The killed random walk on V then corresponds to the transient part of the (full) random walk on $V \cup \bar{a}$ with cemetery at $\bar{a}$. Let $X_0, X_1, X_2, \ldots$ denote the killed random walk on V with

$$\mathbf{P}(X_t = a | X_{t-1} = a') := P_{a'a}$$

for any t. By the Monotone Convergence theorem for conditional expectations we can take the infinite sum inside the expectation in the computation below, and the quantity

$$G_{a'a} := \sum_{t=0}^{\infty} (P^t)_{a'a} = \mathbf{E}\left[\sum_{t=0}^{\infty} I_{X_t=a} \middle| X_0 = a' \right]$$

is the mean expected number of times that the killed random walk started at site a' visits site a, also known as the *Green function* of the random walk. Hence, modulo a multiplicative factor, if $i = (a', a'') \in E$ with $a', a'' \in V$, the identity in Lemma 4.3 represents the difference in the expected number of times the (killed) Markov chain $(X_t)_{t\geq 0}$ visits site a when it starts respectively from $X_0 = a'$ and from $X_0 = a''$.

V. Decay of correlation

We can characterize the identity in Theorem 4.3 in terms of the spectral properties of the transition matrix P, as the following lemma attests. In fact, this lemma considers the matrix $\Gamma := D^{1/2}PD^{-1/2} = D^{-1/2}WD^{-1/2} \in \mathbb{R}^{V\times V}$ that is symmetric and so it is easier to analyze than P. In what follows, let d be the natural distance between vertices on the undirected graph $(V, |E \setminus \partial\bar{a}|)$. Henceforth, we also adopt the convention that $d(\bar{a}, a) = d(a, \bar{a}) = +\infty$ for any $a \in V$.

Lemma 5.1: Let $(\lambda_a, \psi_a)_{a\in V}$ be the pairs of real eigenvalues and orthonormal eigenvectors of the symmetric matrix Γ For each $a \in V$, define the vector $\phi_a \in \mathbb{R}^V$ as $(\phi_a)_c := \sqrt{\sum_{t=0}^{\infty} \lambda_c^t} \frac{1}{\sqrt{d_a}} (\psi_c)_a$, $c \in V$. Then, for each $a \in V$ and $i = (a', a'') \in E$, we have

$$\frac{\partial x_i^\star}{\partial b_a} = w_i (\phi_{a'} - \phi_{a''})^T \phi_a, \tag{4}$$

where we adopt the convention that $\phi_{\bar{a}} := 0$. Moreover, for each $a, a' \in V$ we have

$$\phi_a^T \phi_{a'} = \frac{\mathbf{1}_{a=a'}}{d_{a'}} + \sum_{c\in V} P_{ac} \phi_c^T \phi_{a'}, \tag{5}$$

$$0 \leq \phi_a^T \phi_{a'} \leq \frac{1}{\sqrt{d_a d_{a'}}} \frac{\rho^{d(a,a')}}{1 - \rho}, \tag{6}$$

where $\rho < 1$ is the spectral radius of P.

Proof For simplicity of notation, let us label the elements of V as $\{1, \ldots, p\}$, where $p := |V|$. As $\Gamma := D^{-1/2}WD^{-1/2}$ is real and symmetric, let $(\lambda_\ell, \psi_\ell)_{\ell=1}^p$ be the pairs of real eigenvalues and orthonormal eigenvectors, i.e., $\psi_i^T \psi_j = I_{ij}$, I being the identity matrix. The matrix Γ admits the spectral decomposition $\Gamma = \Psi\Lambda\Psi^T$, where $\Psi = (\psi_1, \ldots, \psi_p) \in \mathbb{R}^{p\times p}$ is orthonormal, $\Psi^T = \Psi^{-1}$, and $\Lambda \in \mathbb{R}^{p\times p}$ is diagonal with entries the eigenvalues of Γ. Clearly, $P = D^{-1/2}\Gamma D^{1/2}$ so that P has eigenvalues/eigenvectors pairs given by $(\lambda_\ell, D^{-1/2}\psi_\ell)_{\ell=1}^p$ (note that the eigenvectors of P are no longer necessarily orthonormal as $(D^{-1/2}\psi_i)^T D^{-1/2}\psi_j = \psi_i D^{-1}\psi_j$) and

$$P^t = D^{-1/2}\Psi\Lambda^t\Psi^T D^{1/2} = D^{-1/2} \left(\sum_{\ell=1}^{p} \lambda_\ell^t \psi_\ell \psi_\ell^T \right) D^{1/2}.$$

As P is sub-stochastic and irreducible, by Lemma 4.2 the spectral radius of P is strictly less than 1, i.e., $\rho < 1$, so that the Neumann series $\sum_{t=0}^{\infty} P^t$ converges. We can write

$$G := \sum_{t=0}^{\infty} P^t = D^{-1/2}\Psi \sum_{t=0}^{\infty} \Lambda^t \Psi^T D^{-1/2} D = \Phi^T \Phi D,$$

where $\Phi := \sqrt{\sum_{t=0}^{\infty} \Lambda^t} \Psi^T D^{-1/2}$. Let $\Phi = (\phi_1, \dots, \phi_p)$ with $\phi_\ell := \Phi e_\ell$ we have $(\phi_\ell)_i = \Phi_{i\ell} = \sqrt{\sum_{t=0}^{\infty} \lambda_i^t} \frac{1}{\sqrt{d_\ell}} (\psi_i)_\ell$ and $G_{ik} = d_k \phi_i^T \phi_k = \sqrt{\frac{d_k}{d_i}} \sum_{\ell=1}^{p} (\psi_\ell)_i (\psi_\ell)_k \sum_{t=0}^{\infty} \lambda_\ell^t$. Clearly, (4) follows from Lemma 4.3. Using the fact that $|\lambda_\ell| \le \rho$ for each ℓ, we get the following bound

$$G_{ik} = \sum_{t=d(i,k)}^{\infty} (P^t)_{ik} \le \sqrt{\frac{d_k}{d_i}} \frac{\rho^{d(i,k)}}{1-\rho} \sum_{\ell=1}^{p} |\psi_\ell|_i |\psi_\ell|_k,$$

where $|\psi_\ell|$ is the vector made by the absolute values of the components of ψ_ℓ. By Cauchy-Schwarz and the orthonormality of Ψ, we have

$$\sum_{\ell=1}^{p} |\psi_\ell|_i |\psi_\ell|_k \le \sqrt{(\Psi\Psi^T)_{ii}} \sqrt{(\Psi\Psi^T)_{kk}} = 1,$$

so that, for each i, k, $0 \le \phi_i^T \phi_k \le \frac{1}{\sqrt{d_i d_k}} \frac{\rho^{d(i,k)}}{1-\rho}$, where the lower bound follows clearly from the fact that $G_{ik} \ge 0$. A first step analysis yields

$$G_{ik} = \sum_{t=0}^{\infty} (P^t)_{ik} = I_{ik} + \sum_{t=1}^{\infty} (P^t)_{ik} = I_{ik} + \sum_{l=1}^{p} P_{i\ell} G_{\ell k}$$

so that $\phi_i^T \phi_k = \frac{\mathbf{1}_{i=k}}{d_k} + \sum_{\ell=1}^{p} P_{i\ell} \phi_\ell^T \phi_k$.

As a corollary of the previous lemma we immediately have the following result, which shows how the optimal network flow problem structurally exhibits exponentially-decreasing correlation bounds with rate given by the spectral radius of the associated sub-stochastic matrix.[2] We use the notation $x \wedge y := \min\{x, y\}$ and $x \vee y := \max\{x, y\}$.

Lemma 5.2 (Point-to-point): Let $\rho < 1$ be the spectral radius of the matrix $\Gamma = D^{1/2} P D^{-1/2}$. For $a \in V$, $i = (a', a'') \in E$,

$$\left| \frac{\partial x_i^\star}{\partial b_a} \right| \le \frac{w_i}{d_a} \mathbf{1}_{\{a'=a \text{ or } a''=a\}} + \gamma \frac{\rho^{(d(a',a) \wedge d(a'',a) - 1) \vee 0}}{1-\rho},$$

where $\alpha := \max_{a \in V, a', a'' \in (V \cup \bar{a})} \frac{\bar{W}_{a'a''}}{d_a}$ and

$$\gamma := \alpha \sum_{c \in V} |P_{a'c} - P_{a''c}| \frac{1}{\sqrt{d_c}},$$

with $P_{\bar{a}a} := 0$ and $d(\bar{a}, a) = +\infty$ for any $a \in V$.

Proof Let $a \in V$ and $i = (a', a'') \in E$. From (4) and (5),

$$\frac{\partial x_i^\star}{\partial b_a} = \frac{w_i}{d_a} (I_{a'a} - I_{a''a}) + w_i \sum_{c \in V} (P_{a'c} - P_{a''c}) \phi_c^T \phi_a,$$

where we adopt the convention that $I_{\bar{a}a} := 0$ and $P_{\bar{a}a} := 0$ for any $a \in V$. Using (6),

$$\left| \frac{\partial x_i^\star}{\partial b_a} \right| \le \frac{w_i}{d_a} |I_{a'a} - I_{a''a}| + \frac{w_i}{\sqrt{d_a}} \frac{1}{1-\rho} \sum_{c \in V} |P_{a'c} - P_{a''c}| \max_{c \in V : P_{a'c} \ne 0 \text{ or } P_{a''c} \ne 0} \frac{\rho^{d(c,a)}}{\sqrt{d_c}}.$$

The statement of the corollary follows immediately from

$$\min_{\substack{c \in V: \\ P_{a'c} \ne 0 \text{ or } P_{a''c} \ne 0}} d(c,a) = \min_{\substack{c \in V: \\ P_{a'c} \ne 0}} d(c,a) \wedge \min_{\substack{c \in V: \\ P_{a''c} \ne 0}} d(c,a) \ge (d(a',a) \wedge d(a'',a) - 1) \vee 0,$$

where we used that, by the triangle inequality for the distance d, (analogously for a'')

$$\min_{\substack{c \in V: \\ P_{a'c} \ne 0}} d(c,a) \ge \min_{\substack{c \in V: \\ P_{a'c} \ne 0}} \{d(a,a') - d(c,a')\} = d(a,a') - 1,$$

with the convention that $\min\{\varnothing\} = +\infty$ (note that $\{c \in V : P_{\bar{a}c} \ne 0\} = \varnothing$), and so also $d(\bar{a}, a) = +\infty$ for any $a \in V$.

Lemma 5.2 yields *point-to-point* correlation bounds as it shows that for each edge i and vertex a the quantity $|\frac{\partial x_i^\star}{\partial b_a}|$ is bounded by a term that decreases exponentially with the distance between i and a. While this result can be used to bound the effect that a perturbation of a *single* component of the constraint vector b has on a *single* component of the optimal solution $x^\star$, this bound is typically not well-suited to capture the *aggregate* impact that *multiple* perturbations have on *multiple* components of the optimal point. Lemma 5.3 below addresses this issue by yielding *point-to-set* and *set-to-point* correlation bounds. The point-to-set bound shows that perturbing *all* the components of the constraint vector b outside a ball of radius r centered at edge i has an effect on $x_i^\star$ that decays exponentially with r. On the other hand, the set-to-point bound shows that perturbing a single component of the constraint vector b at vertex a has an impact on *all* the components of $x^\star$ outside a ball of radius r centered at a that decays exponentially with r. The key property is that these bounds do not depend, respectively, on the number of the components of b being perturbed, nor on the number of the components of $x^\star$ being affected, which would otherwise be the case if we were to use the point-to-point bound in Lemma 5.2. The price to pay for this added versatility is that the bounds in Lemma 5.3 depend on both ρ and ξ, respectively the largest eigenvalue and the corresponding eigenvector of the matrix $\Gamma = D^{1/2} P D^{-1/2}$. On the other hand, the bound in Lemma 5.2 depends only on ρ.

Lemma 5.3 (Point-to-set and set-to-point): Let $\rho < 1$ be the spectral radius of the matrix $\Gamma = D^{1/2} P D^{-1/2}$, and $\xi \in \mathbb{R}^V$ be the corresponding eigenvector. For each $r \ge 1$,

$$\sum_{\substack{a \in V: \\ d(a',a) \wedge d(a'',a) \ge r}} \left| \frac{\partial x_i^\star}{\partial b_a} \right| \le \sigma \frac{\rho^{r-1}}{1-\rho} \qquad \text{for } i = (a', a'') \in E,$$

$$\sum_{\substack{i=(a',a'') \in E: \\ d(a',a) \wedge d(a'',a) \ge r}} \left| \frac{\partial x_i^\star}{\partial b_a} \right| \le \mu \frac{\rho^{r-1}}{1-\rho} \qquad \text{for } a \in V,$$

[2] Lemma 5.2 yields decay of correlation bounds upon the assumption that the spectral radius $\rho < 1$ is not too close to 1, as function of the network size. The behavior of ρ is intrinsically linked to the topology of the underlying graph, and needs to be checked on a case-by-case basis.

where $\alpha := \max_{a\in V, a',a''\in(V\cup\bar{a})} \frac{\bar{W}_{a'a''}}{d_a}$ and

$$\sigma := \alpha\left(\max_{a,c\in V}\frac{\sqrt{d_a}\xi_c}{\sqrt{d_c}\xi_a}\right)\left(\max_{(a',a'')\in E}\sum_{c\in V}|P_{a'c}-P_{a''c}|\right),$$

$$\mu := \alpha\left(\max_{a,c\in V}\frac{\sqrt{d_a}\xi_a}{\sqrt{d_c}\xi_c}\right)\left(\max_{c\in V}\sum_{(a',a'')\in E}|P_{a'c}-P_{a''c}|\right),$$

with $P_{\bar{a}a} := 0$ and $d(\bar{a}, a) = +\infty$ for any $a \in V$.

Proof We first prove the point-to-set bound. Let us define the weighted supremum norm with weight $\omega \in \mathbb{R}^V_+$ as $\|y\|^\omega_\infty := \max_{a\in V}\frac{|y_a|}{\omega_a}$, for any vector $y \in \mathbb{R}^V$. The induced operator norm reads $\|Y\|^\omega_\infty := \max_{a\in V}\frac{1}{\omega_a}\sum_{a'\in V}|Y_{aa'}|\omega_{a'}$, for any matrix $Y \in \mathbb{R}^{V\times V}$. As (ρ, ξ) is an eigenvalue/eigenvector pair for $\Gamma = D^{1/2}PD^{-1/2}$, by defining $\omega := D^{-1/2}\xi$ we have

$$\Gamma\xi = \rho\xi \;\leftrightarrow\; P\omega = \rho\omega \;\leftrightarrow\; \sum_{a'\in V}P_{aa'}\omega_{a'} = \rho\omega_a, a \in V.$$

By the Perron-Frobenius theorem, ξ has strictly positive entries so ω is a well-defined weight vector and clearly $\|P\|^\omega_\infty = \rho$. From the first step analysis

$$G_{a'a} := \sum_{t=0}^{\infty}(P^t)_{a'a} = I_{a'a}+\sum_{t=1}^{\infty}(P^t)_{a'a} = I_{a'a}+\sum_{c\in V}P_{a'c}G_{ca},$$

Lemma 4.3 yields, for $i = (a', a'') \in E$,

$$\frac{\partial x^\star_i}{\partial b_a} = \frac{w_i}{d_a}(I_{a'a} - I_{a''a}) + \frac{w_i}{d_a}\sum_{c\in V}(P_{a'c} - P_{a''c})G_{ca}. \quad (7)$$

Hence, for any $r \geq 1$ we have

$$\sum_{\substack{a\in V:\\ d(a',a)\wedge d(a'',a)\geq r}}\left|\frac{\partial x^\star_i}{\partial b_a}\right| \leq \frac{w_i}{\min_{a\in V} d_a}\sum_{c\in V}|P_{a'c}-P_{a''c}|\cdot$$

$$\cdot \max_{\substack{c\in V:\\ P_{a'c}\neq 0 \text{ or } P_{a''c}\neq 0}} \sum_{\substack{a\in V:\\ d(a',a)\wedge d(a'',a)\geq r}} G_{ca}.$$

For any $c \in V$ such that $P_{a'c} \neq 0$ or $P_{a''c} \neq 0$, by the triangle inequality for the distance d we have

$$d(a', a) \wedge d(a'', a) \leq (d(c, a) + d(a', c)) \wedge (d(c, a) + d(a'', c)),$$

so that $d(a', a) \wedge d(a'', a) \leq d(c, a) + 1$, and since $G_{ca} = \sum_{t=d(c,a)}^{\infty}(P^t)_{ca}$, we get

$$\sum_{\substack{a\in V:\\ d(a',a)\wedge d(a'',a)\geq r}} G_{ca} \leq \sum_{a\in V:d(c,a)\geq r-1} G_{ca}$$

$$\leq \sum_{t=r-1}^{\infty}\sum_{a\in V}(P^t)_{ca} \leq \left(\max_{a,c\in V}\frac{\omega_c}{\omega_a}\right)\sum_{t=r-1}^{\infty}\|P^t\|^\omega_\infty$$

$$\leq \left(\max_{a,c\in V}\frac{\omega_c}{\omega_a}\right)\sum_{t=r-1}^{\infty}(\|P\|^\omega_\infty)^t = \left(\max_{a,c\in V}\frac{\omega_c}{\omega_a}\right)\frac{\rho^{r-1}}{1-\rho}.$$

Combining everything together we clearly get the first bound.

The proof of the set-to-point bound goes analogously. Let us define the weighted $L1$ norm with weight $\omega \in \mathbb{R}^V_+$ as $\|y\|^\omega_1 := \sum_{a\in V}\omega_a|y_a|$, for any vector $y \in \mathbb{R}^V$. The induced operator norm reads $\|Y\|^\omega_1 := \max_{a'\in V}\frac{1}{\omega_{a'}}\sum_{a\in V}\omega_a|Y_{aa'}|$, for any matrix $Y \in \mathbb{R}^{V\times V}$. As Γ is clearly symmetric, by defining $\omega := D^{1/2}\xi$ this time we have

$$\xi^T\Gamma = \rho\xi^T \;\leftrightarrow\; \omega^TP = \rho\omega^T \;\leftrightarrow\; \sum_{a\in V}\omega_aP_{aa'} = \rho\omega_{a'}, a' \in V,$$

and clearly $\|P\|^\omega_1 = \rho$. From identity (7), for any $r \geq 1$,

$$\sum_{\substack{i=(a',a'')\in E:\\ d(a',a)\wedge d(a'',a)\geq r}}\left|\frac{\partial x^\star_i}{\partial b_a}\right| \leq \frac{\max_{i\in E} w_i}{d_a}\sum_{c\in V}\sum_{i=(a',a'')\in E}$$

$$|P_{a'c}-P_{a''c}|G_{ca}\mathbf{1}_{d(a',a)\wedge d(a'',a)\geq r}\mathbf{1}_{P_{a'c}\neq 0 \text{ or } P_{a''c}\neq 0}.$$

As $\mathbf{1}_{d(a',a)\wedge d(a'',a)\geq r}\mathbf{1}_{P_{a'c}\neq 0 \text{ or } P_{a''c}\neq 0} \leq \mathbf{1}_{d(c,a)\geq r-1}$ and

$$\sum_{\substack{i=(a',a'')\in E:\\ d(a',a)\wedge d(a'',a)\geq r}}\left|\frac{\partial x^\star_i}{\partial b_a}\right| \leq \frac{\max_{i\in E} w_i}{d_a}\cdot$$

$$\cdot\left(\max_{c\in V}\sum_{i=(a',a'')\in E}|P_{a'c} - P_{a''c}|\right)\sum_{c\in V:d(c,a)\geq r-1}G_{ca}.$$

Since $G_{ca} = \sum_{t=d(c,a)}^{\infty}(P^t)_{ca}$, we get

$$\sum_{c\in V:d(c,a)\geq r-1}G_{ca} \leq \sum_{t=r-1}^{\infty}\sum_{c\in V:d(c,a)\geq r-1}(P^t)_{ca}$$

$$\leq \sum_{t=r-1}^{\infty}\sum_{c\in V}(P^t)_{ca} \leq \left(\max_{a,c\in V}\frac{\omega_a}{\omega_c}\right)\sum_{t=r-1}^{\infty}\|P^t\|^\omega_1$$

$$\leq \left(\max_{a,c\in V}\frac{\omega_a}{\omega_c}\right)\sum_{t=r-1}^{\infty}(\|P\|^\omega_1)^t = \left(\max_{a,c\in V}\frac{\omega_a}{\omega_c}\right)\frac{\rho^{r-1}}{1-\rho}.$$

Combining everything together we get the second bound.

References

[1] D. Gamarnik, D. Shah, and Y. Wei, "Belief propagation for min-cost network flow: Convergence and correctness," *Operations Research*, vol. 60, no. 2, pp. 410–428, 2012.

[2] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*. Athena Scientific, 1997.

[3] R. L. Dobrušin, "Definition of a system of random variables by means of conditional distributions," *Teor. Verojatnost. i Primenen.*, vol. 15, pp. 469–497, 1970.

[4] B. Simon, *The statistical mechanics of lattice gases. Vol. I*, ser. Princeton Series in Physics. Princeton, NJ: Princeton University Press, 1993.

[5] H.-O. Georgii, *Gibbs measures and phase transitions*, 2nd ed., ser. de Gruyter Studies in Mathematics. Berlin: Walter de Gruyter & Co., 2011, vol. 9.

[6] D. Gamarnik, *Correlation Decay Method for Decision, Optimization, and Inference in Large-Scale Networks*, 2013, ch. 7, pp. 108–121.

[7] C. C. Moallemi and B. Van Roy, "Convergence of min-sum message-passing for convex optimization," *Information Theory, IEEE Transactions on*, vol. 56, no. 4, pp. 2041–2050, 2010.

[8] P. Rebeschini and R. van Handel, "Comparison theorems for Gibbs measures," *Journal of Statistical Physics*, vol. 157, no. 2, pp. 234–281, 2014.

[9] F. Wu and C. Desoer, "Global inverse function theorem," *IEEE Transactions on Circuit Theory*, vol. 19, pp. 199–201, 1972.

[10] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge University Press, 2004.

[11] R. A. Horn and C. R. Johnson, Eds., *Matrix Analysis*. New York, NY, USA: Cambridge University Press, 1986.