Automatica 50 (2014) 910–914

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Brief paper

A stochastic Kaczmarz algorithm for network tomography☆

Gugan Thoppe[a,1], Vivek Borkar[b], D. Manjunath[b]

[a] School of Technology and Computer Science, Tata Institute of Fundamental Research, Mumbai 400005, India

[b] Department of Electrical Engineering, Indian Institute of Technology, Powai, Mumbai 400076, India

ARTICLE INFO

Article history:
Received 17 December 2012
Received in revised form
17 October 2013
Accepted 25 November 2013
Available online 30 December 2013

Keywords:
Kaczmarz algorithm
Stochastic approximation
Network tomography
Online algorithm

ABSTRACT

We develop a stochastic approximation version of the classical Kaczmarz algorithm that is incremental in nature and takes as input noisy real time data. Our analysis shows that with probability one it mimics the behavior of the original scheme: starting from the same initial point, our algorithm and the corresponding deterministic Kaczmarz algorithm converge to precisely the same point. The motivation for this work comes from network tomography where network parameters are to be estimated based upon end-to-end measurements. Numerical examples via Matlab based simulations demonstrate the efficacy of the algorithm.

1. Introduction

1.1. Kaczmarz algorithm

The Kaczmarz algorithm (Kaczmarz, 1937) is a successive projection based iterative scheme for solving ill posed linear systems of equations. Since its introduction, its convergence properties have been extensively analyzed (Galántai, 2004) and it has found diverse applications in areas ranging from tomography (Popa & Zdunek, 2004), synchronization in sensor networks (Freris & Zouzias, 2012), to learning and adaptive control (Åström, 1983; Parks & Militzer, 1992; Richalet, Rault, Testud, & Papon, 1978). The original algorithm is deterministic, but some applications, notably network tomography which we describe later, call for a stochastic version. In this article, we introduce and analyze a stochastic approximation version based on the Robbins–Monro paradigm (Robbins & Monro, 1951) that has become a standard workhorse of signal processing and learning control (Benveniste, Métivier, & Priouret, 1990; Kushner & Yin, 2003). We use the 'o.d.e.' approach (Derevitskii & Fradkov, 1974; Ljung, 1977) to analyze the scheme and argue that it has the same asymptotic behavior as the original deterministic scheme 'almost surely'. While we apply our results to network tomography in this article, we believe that this analysis will be of use in other areas mentioned above. In particular, networked control is one potential application area.

A significant development in this line of research is the randomized Kaczmarz scheme with provable strong convergence properties (Leventhal & Lewis, 2010; Strohmer & Vershynin, 2009), and recent modifications to improve performance by weighted sampling (Freris & Zouzias, 2012; Zouzias & Freris, 2012). The important difference between these works and ours is as follows. For them, the randomization is over the choice of rows, which is a part of the algorithm design and can be chosen at will. In our case, however, a part of the randomness is due to noise and not under our control, as also in the choice of rows which a priori we allow to be uncontrolled.

1.2. Network tomography

Network tomography is the inference of spatially localized network behavior using only measurements of end-to-end aggregates. Recent work can be classified into traffic volume and link delay tomography. A basic paradigm in both these is to infer the statistics of the random vector X from an ill posed measurement model $Y = AX$, where the matrix A is assumed to be known a priori. See Castro, Coates, Liang, Nowak, and Yu (2004), Coates, Hero, Nowak, and Yu (2002) and Lawrence, Michailidis, and Nair (2007) for excellent surveys.

In the transportation literature, the aim is to estimate the traffic volume on the end-to-end routes assuming access to only traffic volumes on a subset of links (Bell, 1991; Maher, 1983; Sherali, Sivanandan, & Hobeika, 1994; Vardi, 1996). An excellent survey

☆ The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Hideaki Ishii under the direction of Editor Ian R. Petersen.

E-mail addresses: gugan@tcs.tifr.res.in (G. Thoppe), borkar.vs@gmail.com (V. Borkar), dmanju@ee.iitb.ac.in (D. Manjunath).

[1] Tel.: +91 22 22782930; fax: +91 22 22282299.

0005-1098/$ – see front matter
http://dx.doi.org/10.1016/j.automatica.2013.12.016

is given in Abrahamsson (1988). An analogous problem has been addressed in packet networking (Feldmann et al., 2001; Zhang, Roughan, Duffield, & Greenberg, 2003; Zhang, Roughan, Lund, & Donoho, 2003). In all of these works, one sample of Y is assumed available and X is estimated by a suitable regularization.

Link delay tomography deals with estimation of link delay statistics from path delay measurements. Here the network is usually assumed to be in the form of a tree. Multicast probe packets, real or emulated, are sent from the root node to the leaves. For each probe packet, a set of delay measurements for paths from the root node to the leaf nodes is collected. These delays are correlated and this correlation is exploited to estimate the link delay statistics (Adams et al., 2000; Cáceres, Duffield, Horowitz, & Towsley, 1999; Duffield, Horowitz, Presti, & Towsley, 2001; Presti, Duffield, Horowitz, & Towsley, 2002). Using T independent samples of the path delay vector, an expectation maximization based algorithm is derived to obtain the maximum likelihood estimates of the link parameters. There is also work on estimating link level loss statistics (Cáceres et al., 1999), link level bandwidths (Downey, 1999), link-level cross traffic (Prasad, Dovrolis, Murray, & Claffy, 2003) and network topology (Duffield, Horowitz, Presti, & Towsley, 2002) using end-to-end measurements.

1.3. Summary of our work

We show that, starting from the same initial point, the stochastic approximation variant of the Kaczmarz (SAK) algorithm and the deterministic Kaczmarz algorithm converge to the same point. Using this, we develop a novel online algorithm for estimation of the means (more generally, moments and cross-moments) of the elements of the vector X from a sequence of measurements of the elements of the vector $Y = AX$. Our scheme can be used for both traffic volume and link delay tomography. An important advantage of our scheme is that it is real time—taking observed data as inputs as they arrive and making incremental adaptation. Also, unlike previous approaches, our scheme allows for elements of X to be correlated. While our analysis is under the simplifying statistical assumption that the samples are IID, we point out in Section 5 that these can be relaxed considerably. For link delay tomography, our algorithm does away with the need for multicast probe packet measurements and can be used even for networks with topologies other than tree.

2. Model and problem description

2.1. Basic notation

For $n \in \mathbb{N}$, $[n] := \{1, \ldots, n\}$. For vectors, we use $\|\cdot\|$ to denote their Euclidean norm and $\langle \cdot, \cdot \rangle$ for inner product. For a matrix A, a_i denotes its ith row, a_{ij} its (i, j)th entry, $\mathcal{R}_A$ its row space and A' its transpose. We use $\dot{x}(t)$ to denote the derivative of the map x with respect to t.

Let $X \equiv (X(1), \ldots, X(N))'$ denote the random vector with finite variance whose statistics we wish to estimate. Let $A \in \mathbb{R}^{m\times N}$, $m < N$, be an a priori known matrix with full row rank and let

$$Y \equiv (Y(1), \ldots, Y(m))' = AX + W, \tag{1}$$

where W is a zero mean, bounded variance random variable denoting noise in the measurement. Let Z be a random variable taking values in $[m]$ such that, $\forall i \in [m]$, $\Pr\{Z = i\} =: \lambda_i > 0$. Let $\{X_k\}, \{Z_k\}, \{W_k\}$, $k \geq 1$, be IID copies of X, Z, W, that are jointly independent, and $Y_k := AX_k + W_k$. (The IID assumption is purely for simplicity of analysis. We point out later that these results extend to much more general situations.) We assume that at each time step k, we know only the value of Z_{k+1} and the Z_{k+1}th component of Y_{k+1}, i.e., $Y_{k+1}(Z_{k+1}) =: \mathcal{Y}_{k+1}$.

Our objective is to develop a real-time algorithm, with provable convergence properties, to estimate the moments and cross-moments of the random vector X.

3. Preliminaries

3.1. Stochastic approximation algorithms

The archetypical stochastic approximation algorithm is

$$x_{k+1} = x_k + \eta_k[h(x_k) + \xi_{k+1}], \tag{2}$$

where $h : \mathbb{R}^n \to \mathbb{R}^n$ is Lipschitz, $\{\eta_k\}_{k\geq 0}$ is a positive stepsize sequence satisfying $\sum_{k\geq 0} \eta_k = \infty$ and $\sum_{k\geq 0}(\eta_k)^2 < \infty$, and ξ_{k+1} represents noise. As $\eta_k \to 0$, (2) can be viewed as a noisy discretization of the o.d.e.

$$\dot{x}(t) = h(x(t)). \tag{3}$$

This is the 'o.d.e. approach' (Derevitskii & Fradkov, 1974; Ljung, 1977). More specifically, suppose that the following assumptions hold.

(A1) $\{\xi_k\}$ is a square-integrable martingale difference sequence w.r.t. the σ-fields $\{\mathcal{F}_k\}$, $\mathcal{F}_k := \sigma(x_0, \xi_1, \ldots, \xi_k)$, satisfying $E[\|\xi_{k+1}\|^2|\mathcal{F}_k] \leq L(1 + \|x_k\|^2)$ a.s. for some $L > 0$.

(A2) $\forall u$, $h_\infty(u) := \lim_{c\uparrow\infty} h(cu)/c$ exists (h_∞ will be necessarily Lipschitz) and the o.d.e. $\dot{x}(t) = h_\infty(x(t))$ has origin as its globally asymptotically stable equilibrium.

(A3) $H := \{x \in \mathbb{R}^n : h(x) = 0\} \neq \emptyset$. Also, $\exists$ a continuously differentiable Lyapunov function $\mathcal{L} : \mathbb{R}^n \to \mathbb{R}$ such that $\langle \nabla\mathcal{L}(x), h(x)\rangle < 0$ for $x \notin H$.

Then, as in Chapters 2, 3 of Borkar (2008), we have the following lemma.

Lemma 1. *The iterates $\{x_k\}$ of (2) a.s. converge to H.*

3.2. Kaczmarz algorithm

Consider the inverse problem of finding a fixed $v^* \in \mathbb{R}^N$ from Av^*, where A is as defined in Section 2. W.l.o.g., let rows of A be of unit norm. Given an approximation x_0 of v^*, a natural optimization problem to consider is

$$\min_{u\in\mathbb{R}^N} \|u - x_0\|, \quad \text{subject to } Au = Av^*. \tag{4}$$

Elementary calculation shows that its solution is

$$x^* = x_0 + A'(AA')^{-1}(Av^* - Ax_0). \tag{5}$$

Clearly, $x^* \in \mathcal{A}^0 := x_0 + \mathcal{R}_A$. As A has full row rank, x^* is the only point in $\mathcal{A}^0$ that satisfies $Au = Av^*$. The Kaczmarz algorithm uses this fact to solve (4). With prescribed initial point x_0, stepsize κ, and $r_k \equiv (k \bmod m) + 1$, its update rule is given by

$$x_{k+1} = x_k + \kappa[\langle a_{r_k}, v^*\rangle - \langle a_{r_k}, x_k\rangle]a_{r_k}. \tag{6}$$

Theorem 1 (*Chong & Zak, 2001*). *If $0 < \kappa < 2$, then $x_k \to x^*$ as $k \to \infty$.*

Let $\mathcal{A}^* := v^* + \mathcal{R}_A$. Since $\mathcal{A}^0$, $\mathcal{A}^*$ are translations of $\mathcal{R}_A$, $\text{dist}(x_0, \mathcal{A}^*) = \text{dist}(\mathcal{A}^0, v^*)$. As $A(x^* - v^*) = 0$, $(x^* - v^*) \perp \mathcal{R}_A$. Thus, $(x^* - v^*) \perp \mathcal{A}^0, \mathcal{A}^*$. Hence, $\|v^* - x^*\| = \text{dist}(\mathcal{A}^0, v^*) = \text{dist}(x_0, \mathcal{A}^*)$. Thus we have the following lemma.

Lemma 2. *For any $\delta > 0$, $\|x^* - v^*\| < \delta$ if and only if $\text{dist}(x_0, \mathcal{A}^*) < \delta$.*

4. The SAK algorithm

We develop here a SAK algorithm to estimate $\mathbb{E}X$ for the model of Section 2. Let x_0, an approximation to $\mathbb{E}X$, be given. Observe from (1) that

$$\mathbb{E}Y = A\mathbb{E}X. \tag{7}$$