Variance Reduction in Stochastic Methods for Large-scale Regularised Least-square Problems

Yusuf Yiğit PİLAVCI, Pierre-Olivier AMBLARD, Simon BARTHELMÉ, Nicolas TREMBLAY

Introduction

- Linear least-square problems are expensive to solve in large dimensions.
- State-of-the-art algorithms consists of approximate methods, both deterministic and stochastic.
- An interesting stochastic approach is based on determinantal point processes (DPPs).
- This approach yields an unbiased estimator of the solution but still might suffer from high variance.

In this work, we propose a simple variance reduction technique for the DPP-based estimator for estimating the solution of the regularized least square problem. We apply this technique on the estimators based on random spanning forests to solve graph Tikhonov regularization.

Problem Definition

Regularized Regression on Graphs

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{z}\in\mathbb{R}^n} q\underbrace{||\mathbf{y}-\mathbf{z}||^2}_{\text{Fidelity}} + \underbrace{\mathbf{z}^T\mathsf{L}\mathbf{z}}_{\text{Regularization}}, \quad q>0$$

where $\mathbf{y}\in\mathbb{R}^n$ is a graph signal. L denotes the graph Laplacian of the given graph and q is the regularization parameter.

The explicit solution to this problem is:

$$\hat{\mathbf{x}} = \mathsf{K}\mathbf{y} \text{ with } \mathsf{K} = (\mathsf{L}+q\mathsf{I})^{-1}q\mathsf{I}$$

where I is the identity matrix.

- Direct computation of K requires $\mathcal{O}(n^3)$ elementary operations due to the inverse.
- For large n, iterative methods and polynomial approximations are state-of-the-art. Both compute $\hat{\mathbf{x}}$ in linear time in the number of edges $|\mathcal{E}|$.

Random Spanning Forest based Estimators

For an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathsf{W})$:

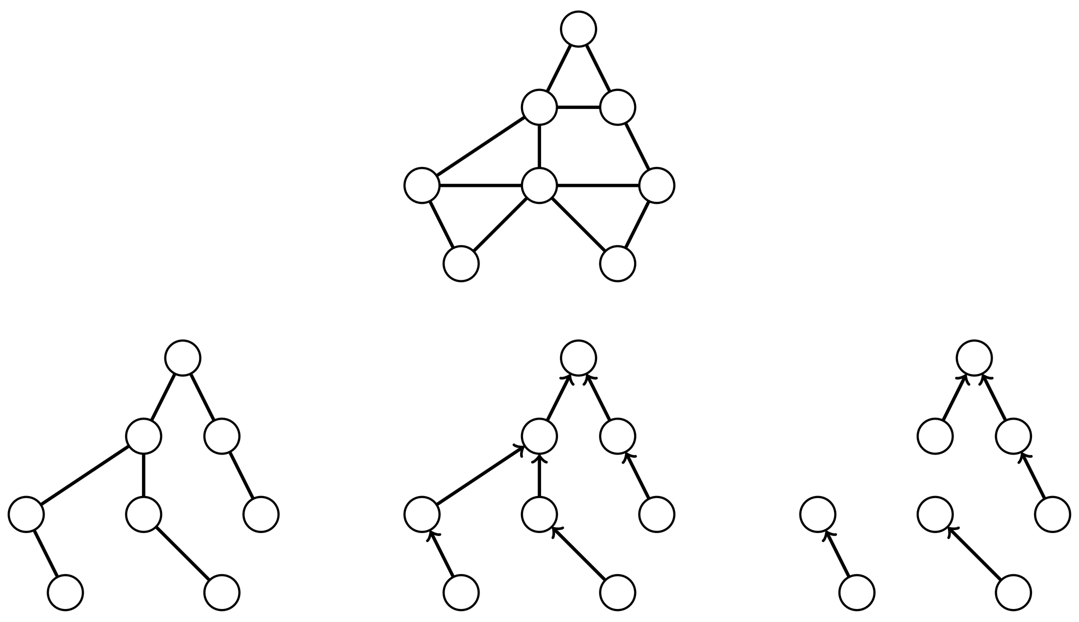

Fig. 1: Original graph, a spanning tree, a rooted spanning tree and a rooted spanning forest

Random Spanning Forests (RSF)

Consider the following parametric distribution over rooted spanning forests:

$$P(\Phi_q=\phi) \propto q^{|\rho(\phi)|}\prod_{\tau\in\phi}\prod_{(i,j)\in\tau}\mathsf{W}_{i,j}$$

where q is a parameter and $\rho(\phi)$ denotes the set of roots in the forest ϕ. One can sample from this distribution by a variant of Wilson's algorithm in time $\mathcal{O}(|\mathcal{E}|/q)$.

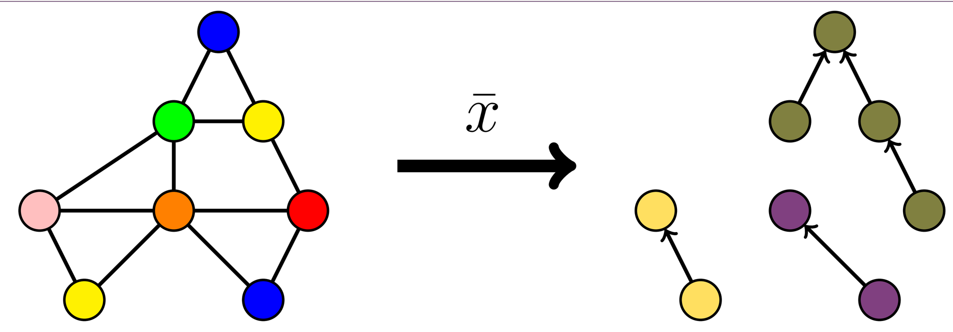

Fig. 2: An illustration for the estimator

A key result: $\bar{\mathbf{x}}$ is and unbiased estimator of $\hat{\mathbf{x}}$ with the expected error $\mathbf{y}^\top(\mathsf{K}-\mathsf{K}^2)\mathbf{y}$.

Proposed Method

$\hat{\mathbf{x}}$ minimizes the following cost function:

$$F(\mathbf{z}) = \frac{1}{2}\mathbf{z}^\top\mathsf{K}^{-1}\mathbf{z} - \mathbf{z}^\top\mathbf{y}$$

Gradient Descent Update as Control Variate

We propose to apply the gradient descent update on the previous estimator $\bar{\mathbf{x}}$:

$$\bar{\mathbf{z}} := \bar{\mathbf{x}} - \alpha(\mathsf{K}^{-1}\bar{\mathbf{x}} - \mathbf{y})$$

Range of α. For certain values of α, $\bar{\mathbf{z}}$ have improved performance, *e.g.* $\alpha = \frac{2q}{2q+d_{max}}$. The optimal value is $\alpha^\star = \frac{\mathrm{tr}(\mathrm{Cov}(\mathsf{K}^{-1}\bar{\mathbf{x}},\bar{\mathbf{x}}))}{\mathrm{tr}(\mathrm{Var}(\mathsf{K}^{-1}\bar{\mathbf{x}}))}$.

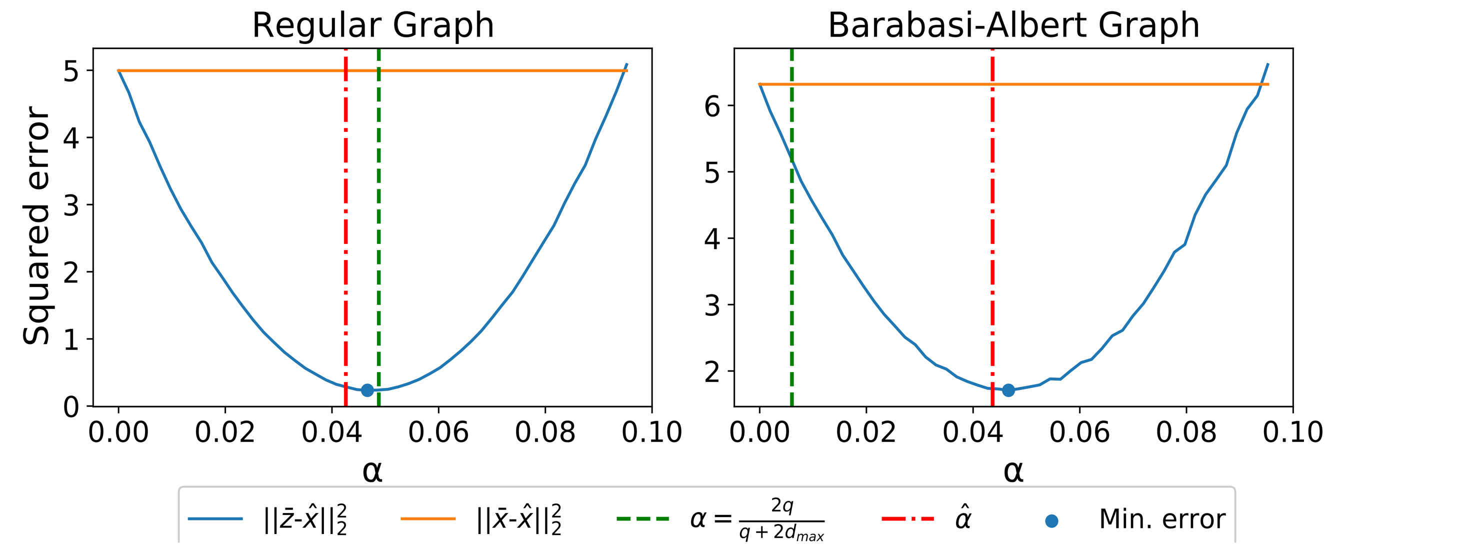

Fig. 3: An empirical comparison of different choices of α

Computational cost. A matrix-vector product with L is needed only once.

Experiments

Denoising graph signals via Tikhonov regularization:

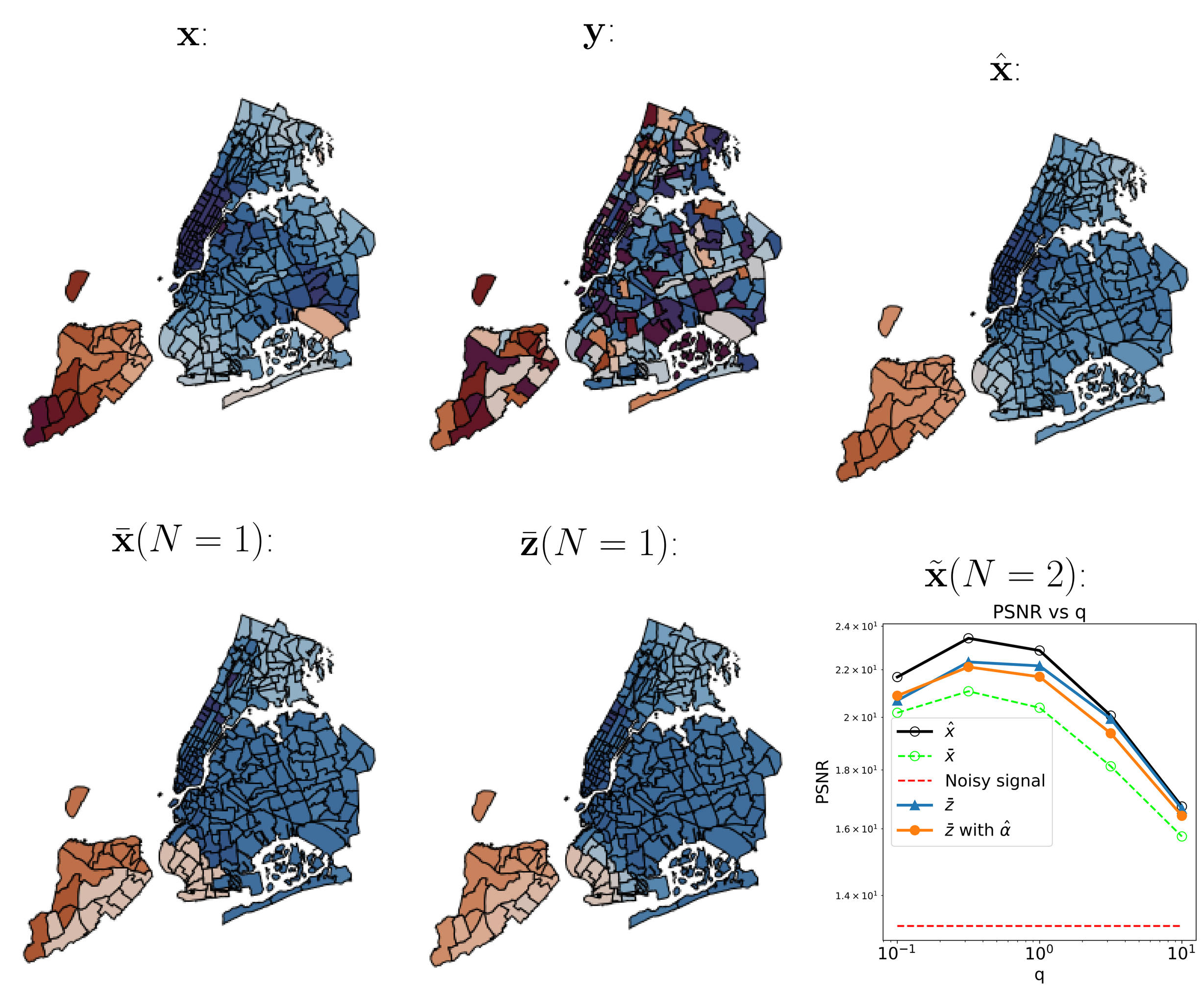

Fig. 4: Graph signal smoothing via Tikhonov regularization illustrated on NYC taxi dataset.

SSL for node classification aims to classify vertices while the class information (or labels) is available over only a few vertices. Given labels in $\mathbf{y}_l$ for class l, the solution (Avrachenkov *et.al.*) under a smoothness prior regulated by $\mu > 0$ is the classification function $\mathbf{f}_l = \mathsf{D}^{1-\sigma}\mathsf{K}\mathsf{D}^{\sigma-1}\mathbf{y}_l$ where $\mathsf{K} = (\mathsf{Q}+\mathsf{L})^{-1}\mathsf{Q}$ and $\mathsf{Q} = \frac{\mu}{2}\mathsf{D}$.

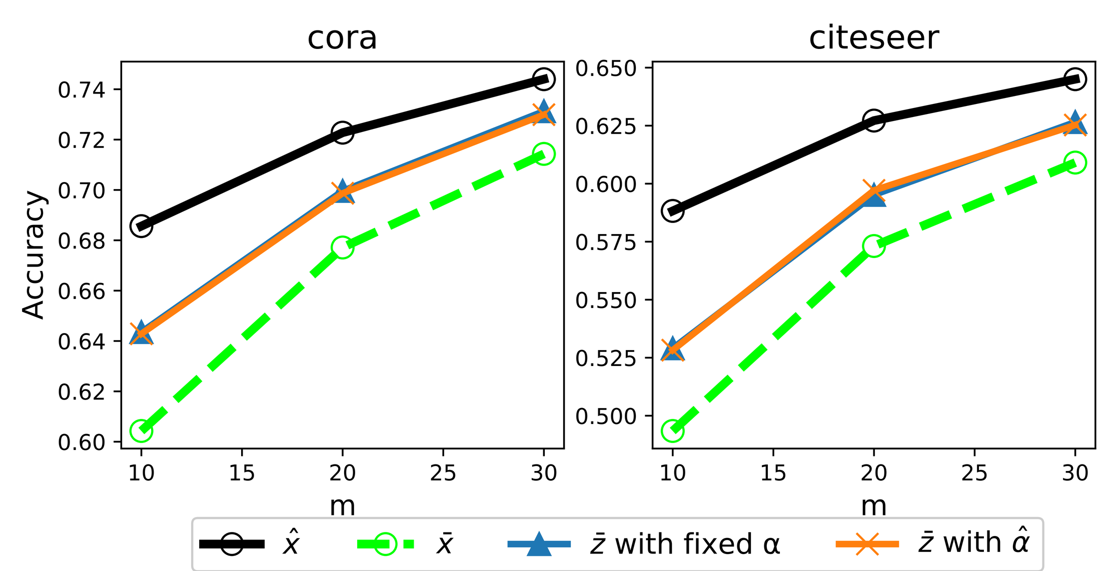

Fig. 5: SSL via the proposed method on Cora and Citeseer citation networks

Grenoble Images Parole Signal Automatique
UMR CNRS 5216 - Grenoble Campus
38400 Saint Martin d'Hères - FRANCE