Dantzig Selector with an Approximately Optimal Denoising Matrix and its Application in Sparse Reinforcement Learning

Bo Liu
Auburn University
boliuauburn@gmail.com

Luwan Zhang
University of Wisconsin-Madison
luwanzhang@gmail.com

Ji Liu
University of Rochester
ji.liu.uwisc@gmail.com

Abstract

Dantzig Selector (DS) is widely used in compressed sensing and sparse learning for feature selection and sparse signal recovery. Since the DS formulation is essentially a linear programming optimization, many existing linear programming solvers can be simply applied for scaling up. The DS formulation can be explained as a basis pursuit denoising problem, wherein the data matrix (or measurement matrix) is employed as the denoising matrix to eliminate the observation noise. However, we notice that the data matrix may not be the optimal denoising matrix, as shown by a simple counter-example. This motivates us to pursue a better denoising matrix for defining a general DS formulation. We first define the optimal denoising matrix through a minimax optimization, which turns out to be an NP-hard problem. To make the problem computationally tractable, we propose a novel algorithm, termed as "Optimal" Denoising Dantzig Selector (ODDS), to approximately estimate the optimal denoising matrix. Empirical experiments validate the proposed method. Finally, a novel sparse reinforcement learning algorithm is formulated by extending the proposed ODDS algorithm to temporal difference learning, and empirical experimental results demonstrate to outperform the conventional "vanilla" DS-TD algorithm.

1 Introduction

We consider consider the classic problem in compressed sensing, sparse learning, and statistics [Donoho, 2006, Candes and Tao, 2007, Bickel et al., 2009]:

Given a data (measurement) matrix $X \in \mathbb{R}^{n\times m}$ ($m \gg n$) and a noisy observation vector $y \in \mathbb{R}^n$ satisfying $y = X\beta^* + \epsilon$ where ϵ is the noise vector following the Gaussian distribution $N(0, \sigma^2 I)$[1] *and β^* is the truth model which is a sparse vector. How to recover the sparse vector β^* from this under-determined system?*

Dantzig Selector (DS) [Candes and Tao, 2007] is a widely used approach to solving this problem. The standard DS is formulated as

$$\text{(DS)} \quad \hat{\beta}_{\text{DS}} = \underset{\beta}{\operatorname{argmin}} \ \|\beta\|_1 \tag{1a}$$

$$\text{s.t. } \|X^T(X\beta - y)\|_\infty \leq \lambda. \tag{1b}$$

DS has a very similar performance to another famous formulation LASSO [Tibshirani, 1996] both empirically and theoretically [Bickel et al., 2009]. The DS formulation (1) can be formulated as a linear programming (LP) problem, thus many matured LP solvers can be directly applied to address this problem with large-scale problem settings. The DS formulation or its variation has been widely used in reinforcement learning [Geist et al., 2012, Liu et al., 2012, Mahadevan and Liu, 2012, Qin et al., 2014], computational bioinformatics [Liu, 2014], and computer vision [Cong et al., 2011].

The motivation of this paper is to explore the role of X^T (the transpose of X) in DS formulation (1). We note that the constraint in (1b) follows two principles: 1) the defined feasible region should contain the true solution β^* with high probability; and 2) to make $\hat{\beta}_{\text{DS}}$ close to β^* the feasible region defined by the constraint should be as small as possible, that is, λ is expected to a small value. Taking $\beta = \beta^*$ into the constraint leads to the smallest possible value for $\lambda = \|X^T(X\beta^* - y)\|_\infty = \|X^T\epsilon\|_\infty$. If columns of X are normalized to 1, we have $\lim_{n\to\infty} \|X^T\epsilon\|_\infty \to 0$ with high probability Candes and Tao [2007]. Therefore, the factor X^T in the constraint actually plays the role of *denoising*. This motivates us to ask two questions: 1) is X^T the optimal denoising matrix for the recovery of the sparse signal β^*?; and 2) if not, how to measure the the optimality of the denoising matrix and how to compute the optimal denoising matrix?

[1] The Gaussian distribution can be generalized to any sub-Gaussian distribution.

Unfortunately, X^T is *not* the optimal denoising matrix in general. We provide a counter-example in Section 3. The main contributions of this paper are summarized below:

- We propose a generalized denoising Dantzig selector formulation (GDDS) and define the optimal denoising matrix for sparse signal recovery via a minimax formulation;
- A two-stage approach is proposed to compute the approximately optimal denoising matrix;
- We apply the proposed ODDS algorithm to an important application in reinforcement learning: the temporal difference learning problem for sparse value function approximation.

This paper is organized as follows: Related work is introduced in Section 2. Section 3, which is the core part of this paper, proposes the general Dantzig Selector formulation with both intuitive motivations as well as the mathematical backgrounds. A generalized error bound is proposed, which leads to the definition of the optimal denoising matrix, which is NP-hard in general. To address this problem, a two-stage algorithm for approximately computing the optimal denoising matrix is given. Then in Section 4, the algorithm is applied to reinforcement learning to design a new algorithm for sparse value function approximation. The experimental results are presented in Section 5, which validate the effectiveness of the proposed algorithm.

2 Related Work

The problem considered in this paper has received substantial attentions in compressed sensing, sparse learning, and statistics. The study starts from the special case, i.e. the noiseless case $\epsilon = 0$. To recover the sparse vector β^*, there are two types of approaches: ℓ_1 norm minimization and greedy algorithms. The ℓ_1 norm minimization solves the following optimization problem to estimate β^* [Chen and Donoho, 1994, Candes and Tao, 2005]

$$\min_{\beta} \quad \|\beta\|_1 \quad \text{s.t.} \quad X\beta = y. \tag{2}$$

The theoretical study [Candes and Tao, 2005] suggests that under the restricted isometric property (RIP) condition, the ℓ_1 norm minimization formulation can recover the true solution β^* exactly. The greedy approaches include the forward greedy algorithm (or OMP) [Tropp, 2004] and the backward greedy algorithm. A similar theoretical guarantee of exact recovery is proven for the forward greedy algorithm in Zhang [2011b].

Now let us turn to the noisy case, i.e. $\epsilon \neq 0$. The noisy case is more challenging than the noiseless case. It also mainly includes two types of approaches: ℓ_1 norm minimization approaches and greedy algorithms. To deal with the noise, there are several popular formulations including DS [Candes and Tao, 2007], LASSO [Tibshirani, 1996], and basis pursuit denoising (BPDN) [Chen et al., 2001]. They essentially apply different manners to denoise. BPDN uses the ℓ_2 norm penalty to denoise:

$$\hat{\beta}_{\text{BPDN}} = \operatorname*{argmin}_{\beta} \|\beta\|_1 \quad \text{s.t.} \quad \|X\beta - y\|_2 \leq \varepsilon, \tag{3}$$

where the constraint basically restricts the noise ϵ by $\|\epsilon\| \leq \varepsilon$. DS uses the ℓ_∞ norm penalty to denoise $\|X^T\epsilon\|_\infty \leq \lambda$ as shown above. LASSO applies the same spirit as DS [Bickel et al., 2009] to denoise, but uses a different formulation

$$\hat{\beta}_{\text{LASSO}} = \operatorname*{argmin}_{\beta} \frac{1}{2}\|X\beta - y\|_2^2 + \lambda\|\beta\|_1. \tag{4}$$

The theoretical error bound for LASSO and DS is similar [Bickel et al., 2009] and better than BPDN in some sense. The key reason lies in that the noise constraint used in DS and LASSO $\|X^T\epsilon\|_\infty \leq \lambda$ is sharper than the noise constraint used in BPDN $\|\epsilon\| \leq \varepsilon$.[2] The greedy approaches mainly include the forward greedy algorithm [Zhang, 2009] and the forward-backward greedy algorithm [Zhang, 2011a, Liu et al., 2013].

While Dantzig Selector primarily focuses on ℓ_1 regularization for sparsity, the group-sparsity structures was explored [Liu et al., 2010a], and was recently extended to generalized norm [Chatterjee et al., 2014]. Other variants includes weighted Dantzig Selector [Candes et al., 2008] by re-weighting the sparse signal, multi-stage Dantzig Selector [Liu et al., 2010b] for iterative sparse signal recovery, etc. As for the computational achievements, besides the primal-dual interior point method [Candes and Tao, 2007], TFOCS [Becker et al., 2011] is also widely used. Later, inexact alternating direction method of multipliers (ADMM) formulations are proposed in [Lu et al., 2012, Wang and Yuan, 2012], which are computationally efficient.

3 Algorithm

We first show the generic error bound when "X^T" in (1b) is substituted by an arbitrary denoising matrix Q^T. Then a counter example is provided to show why X^T is not the optimal choice for Q^T. We prove an approximate method to pursue the optimal denoising matrix Q^T in the end of this section.

3.1 Generalized Denoising Dantzig Selector and its Error Bound

With the backgrounds of the denoising matrix introduced above, one intuitive question is that if X^T the optimal denoising matrix for sparse signal recovery of β^*? The answer is actually NO! To explain this, we first introduce the

[2]From the optimization perspective, BPDN seems consistent with LASSO, but the theoretical error bound is worse than LASSO and DS for some subtle reasons, which is beyond the scope of this paper.

generalized denoising Dantzig Selector formulation. Next, an error bound w.r.t the GDDS and regular DS is proposed, and it can be proven that for regular DS, the error bound is tighter than the existing error bound provided in [Candes and Tao, 2007] and [Bickel et al., 2009]. Then a simple counter-example is proposed to argue that X^T may not be the optimal denoising matrix. We here present some definitions and notations in Figure 1.

- h_T, h_{T^c}: T and T^c are two complementary subsets in $\{1, 2, \cdots, m\}$. We denote by $h_T \in \mathbb{R}^m$ the vector taking the same values as h on T and zeros in the rest; the same for $h_{T^c} \in \mathbb{R}^m$;
- $|T|$ returns the cardinality of the set T;
- $\|W\|_\infty$ is defined as the induced ∞ norm of matrix $W \in \mathbb{R}^{m\times m}$, i.e. $\|W\|_\infty = \max_x \frac{\|Wx\|_\infty}{\|x\|_\infty}$ where the vector infinity norm $\|Wx\|_\infty$ and $\|x\|_\infty$ are defined as usual;
- $\leq_{(\mathrm{P})}$: less than or equal to with high probability;

Figure 1: Notation used in this paper

First we define a generalized denoising Dantzig Selector (GDDS) formulation by using a general denoising matrix Q^T ($Q \in \mathbb{R}^{n\times m}$) to replace the X^T in the constraint:

$$\begin{aligned}(\text{GDDS}) \quad & \min_\beta : \ \|\beta\|_1 \\ & \text{s.t.} : \ \|Q^T(X\beta - y)\|_\infty \leq \lambda.\end{aligned} \tag{5}$$

Next we will propose an error bound for the proposed GDDS in (5) and regular DS. Since the denoising matrix is not necessarily X^T anymore, the commonly used RIP condition or restricted eigenvalue (RE) condition [Van De Geer et al., 2009] are not eligible here. To extend to the general case, we define a new condition termed as *generalized restricted (GR) constant*.

Definition 1. *(GR constant) Given $X, Q \in \mathbb{R}^{n\times m}$ and $p \in [1, \infty]$, the general restricted constant $\rho(Q^TX, p, s)$ is defined as*

$$\rho(Q, X, p, s) := \min_{|T|\leq s, \|h_{T^c}\|_1 \leq \|h_T\|_1} \frac{\|Q^TXh\|_\infty}{\|h\|_p}. \tag{6}$$

This definition essentially provides the lower bound of the ratio between $\|Q^TXh\|_\infty$ and $\|h\|_p$ over h in a subset of $\mathbb{R}^p$, which characterizes the property of Q^TX. The GR constant leads to a weaker condition to exactly recover the sparse signal β^* for the noiseless case ($\epsilon = 0$) and a tighter error bound for the noisy case ($\epsilon \neq 0$) than the existing analysis, as shown by Theorem 1. Based on the definition for GR constant in Definition 1, we have the following error bound on $\|\hat{\beta}_{\text{GDDS}} - \beta^*\|_p$.

Theorem 1. *Assume that the GR condition is satisfied, i.e. the GR constant $\rho(Q, X, p, \|\beta^*\|_0) > 0$. Choose $\lambda = \|Q^T\epsilon\|_\infty$ in (5). We have*

$$\|\hat{\beta}_{GDDS} - \beta^*\|_p \leq \frac{2\|Q^T\epsilon\|_\infty}{\rho(Q, X, p, \|\beta^*\|_0)}, \tag{7}$$

where $\hat{\beta}_{GDDS}$ is the solution to (5) and p can be any value in the range $[1, \infty]$.

It should be noted that albeit with a more general error bound, Theorem 1 does not weaken the existing analysis based on the following two observations:

- In the noiseless case, i.e. $\epsilon = 0$, if the GR condition is satisfied, then $\hat{\beta}_{\text{GDDS}}$ is able to exactly recover β^*. Note that the GR condition is weaker than the RIP condition for $Q = X$ [Candes, 2008]. In other words, the RIP condition leads to GR condition when $Q = X$. The detailed interpretation is provided in Appendix.
- In the noisy case, i.e. $\epsilon \neq 0$, this bound (7) is a tighter bound than the bound $\|\hat{\beta}_{\text{DS}} - \beta^*\|_p$ for DS in [Candes and Tao, 2007, Bickel et al., 2009] for $Q = X$. Please also refer to Appendix for detailed comparisons.

The key reason why Theorem 1 does not loose the existing analysis lies in that the definition of the GR constant $\rho(Q, X, p, s)$ skips many relaxation steps by directly indicating the relationship between $\|Q^T\epsilon\|_\infty$ and $\|\hat{\beta}_{\text{GDDS}} - \beta^*\|^2$. Given p, s and X, $\rho(Q, X, p, s)$ reflects the ability for sparse signal recovery of β^* of the denoising matrix Q^T: *The larger $\rho(Q, X, p, s)$ is, the better Q is able to recover the sparse signal β^*.* Therefore, since the error bound provided in Theorem is tight enough, it is reasonable to use the bound in (7) as the evaluation criteria to see why X^T is not the optimal denoising matrix and define the optimal denoising matrix.

3.2 A Counter-Example

To see why X^T may not be the optimal denoising matrix, we show an example of Q which gives a lower value for the bound in (7). To construct such a matrix Q, we require the same conditions on Q as X, i.e. all columns of Q have been normalized with norm 1. We can verify that any Q with unit column norm, the value of $\|Q^T\epsilon\|_\infty$ is comparable to $\|X^T\epsilon\|_\infty$, according to the following standard results in Lemma 1.

Lemma 1. *For any $Q \in \mathbb{R}^{n\times m}$ satisfying $\|Q_{.i}\| = 1$ for $i = 1, \cdots, m$, we have*

$$\|Q^T\epsilon\|_\infty \leq \sigma\sqrt{\log m}$$

with high probability at least $1 - O(1/m)$.

The key reason why $\|Q^T\epsilon\|_\infty$ and $\|X^T\epsilon\|_\infty$ are comparable lies in that all entries in random vectors $Q^T\epsilon \in \mathbb{R}^p$

and $X^T\epsilon \in \mathbb{R}^p$ follow the same Gaussian distribution $\mathcal{N}(0, \sigma^2)$. Since $\|Q^T\epsilon\|_\infty$ and $\|X^T\epsilon\|_\infty$ are comparable, we only need to find an example for Q such that $\rho(Q, X, p, s) > \rho(X, X, p, s)$ for some X. Let $s = 1$ for simplicity and

$$X = \begin{pmatrix} \sqrt{3}/2 & 1/2 \\ 1/2 & \sqrt{3}/2 \end{pmatrix}, \; Q = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{8}$$

We have

$$X^T X = \begin{pmatrix} 1 & \sqrt{3}/2 \\ \sqrt{3}/2 & 1 \end{pmatrix} \tag{9}$$

$$\begin{aligned} \rho(X, X, p, 1) = & \min_{\substack{|T|\leq 1 \\ \|h_{T^c}\|_1 \leq \|h_T\|_1}} \frac{\|X^T X h\|_\infty}{\|h_T\|_p} \\ = & 1 - \sqrt{3}/2 = 0.134, \\ \rho(Q, X, p, 1) = & \min_{\substack{|T|\leq 1 \\ \|h_{T^c}\|_1 \leq \|h_T\|_1}} \frac{\|Q^T X h\|_\infty}{\|h_T\|_p} \\ = & \sqrt{3}/2 - 1/2 = 0.366, \end{aligned}$$

where in both cases the optimal value of h is $h = [1, -1]^T$ regardless of the value of p. Thus, this example shows that X^T may not always be the optimal denoising matrix.

3.3 Optimal Denoising Matrix and its Approximation

From the counter-example in the previous section, we know that X is not the optimal option for Q to maximize $\rho(Q, X, p, s)$. Therefore, it raises an optimization problem to find the optimal Q:

$$Q^* = \operatorname*{argmax}_{\|Q_{.i}\| \leq 1} \min_{\substack{|T|\leq s \\ \|h_{T^c}\|_1 \leq \|h_T\|_1}} \frac{\|Q^T X h\|_\infty}{\|h_T\|_p}. \tag{10}$$

However, this formulation is extremely difficult to solve. Although we can solve it easily for small n, m as in our example above, it is NP-hard in general. Therefore, it is unrealistic to solve this problem exactly. To find a reasonable approximation, we consider an alternative way to finding an optimal matrix $W \approx Q^T X$:

$$W^* = \operatorname*{argmax}_{\|W\|_\infty \leq 1} \min_{\substack{|T|\leq s \\ \|h_{T^c}\|_1 \leq \|h_T\|_1}} \frac{\|W h\|_\infty}{\|h_T\|_p}. \tag{11}$$

One can easily verify that the optimal solution is $W^* = I$.[3] The second step is to find the optimal Q. We try to find the best Q from another perspective. Intuitively, we want $Q^T X$ to be close to the identity matrix. We estimate Q by the following:

$$\min_Q : \; \|Q^T X - I\|_p \quad \text{s.t.} : \quad \|Q_{.i}\| \leq 1 \tag{12}$$

[3] Actually, it is not important which norm is used to restrict W. For most norms we verified, the optimal solution for W should be a diagonal matrix with equal values.

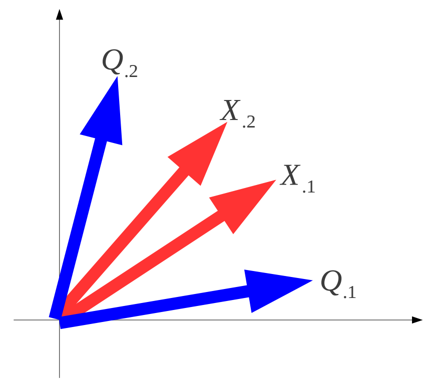

Figure 2: An example of X and Q.

where $p \in [1, \infty]$ and $\|X\|_p := (\sum_{i,j} |X_{ij}|^p)^{1/p}$. We have many options for choosing $p \in [1, \infty]$. Based on our empirical study, a reasonable empirical value for p is $p = 2$, and thus the problem can be recast as a strongly convex problem

$$\min_Q : \; \|Q^T X - I\|_2^2 \quad \text{s.t.} : \quad \|Q_{.i}\| \leq 1 \tag{13}$$

There are many optimization algorithms to address this problem. We use Nesterov's accelerated gradient method [Nesterov, 2004]. Note that the empirical Q obtained from (12) or (13) is generally different from the optimal one defined in (10). And the optimal denoising Dantzig Selector algorithm is summarized as in Algorithm 1.

Algorithm 1 "Optimal" Denoising Dantzig Selector (ODDS)

Require: $X \in \mathbb{R}^{n\times m}, y \in \mathbb{R}^n$
Ensure: β
  Compute the denoising matrix Q^T via Eq. (13)
  Compute β via Eq. (5)

Figure 2 provides an example of Q and X in the two-dimensional case. Intuitively, Q is an approximation to X by slightly modifying all feature (column) vectors of X such that they are as different from each other as possible.

4 Reinforcement Learning

Dantzig selector has an important application in reinforcement learning. This section first briefly introduces reinforcement learning and then shows how to apply the proposed ODDS method to it.

A *Markov Decision Process* (MDP) is defined by the tuple $(S, A, P^a_{ss'}, R, \gamma)$, comprised of a set of states S, a set of actions A, a dynamical system model comprised of the transition kernel $P^a_{ss'}$ specifying the probability of transition from state $s \in S$ to state $s' \in S$ under action $a \in A$, a reward model $R(s, a) : S \times A \to \mathbb{R}$, and $0 \leq \gamma < 1$ is a discount factor. A policy $\pi : S \to A$ is a deterministic mapping from states to actions. Associated with each policy π is a value function V^π, which is the fixed point of the Bellman equation:

$$V^\pi = T^\pi(V^\pi) = R^\pi + \gamma P^\pi V^\pi,$$

for a given state $s \in S$, $R^\pi(s) = R(s, \pi(s))$, and P^π is the state transition function under fixed policy π, and T^π is known as the *Bellman operator*. In what follows, we often drop the dependence of V^π, T^π, R^π on π, for notation simplicity. In linear value function approximation, a value function is assumed to lie in the linear span of a basis function matrix Φ of dimension $|S| \times d$, where d is the number of linear independent features. It is easy to show that the "best" approximation $\hat{v}_{\text{best}}$ of the true value function V satisfies the following equation,

$$\hat{v}_{\text{best}} = \Pi V = \Pi L^{-1} R, \tag{14}$$

where $\Pi = \Phi(\Phi^T \Xi \Phi)^{-1} \Phi^T \Xi$, Ξ is a diagonal matrix where the i-th diagonal entry ξ_i is the stationary state distribution w.r.t state s_i. and $L = I - \gamma P$. However, L^{-1} is also computationally prohibitive. To this end, a more practical way is to compute a Galerkin-Bubnov approximate solution [Yu and Bertsekas, 2010] is via solving a fixed-point equation

$$\hat{v} = \Pi_\Phi^X T \hat{v} \tag{15}$$

w.r.t an oblique projection Π_Φ^X onto $\text{span}(\Phi)$ orthogonal to X, i.e., $\Pi_\Phi^X = \Phi(X^T\Phi)^{-1}X^T$. The existence and uniqueness of the solution can be verified since T is a contraction mapping, Π_Φ^X is a non-expansive mapping, and thus $\Pi_\Phi^X T$ is a contraction mapping. As for the optimal solution $\hat{v}_{\text{best}}$, we have $X_{\text{best}} = (L^T)^{-1}\Xi\Phi$, which is also computational expensive. An often used X used in the fixed-point equation (15) is $X_{TD} = \Xi\Phi$, which is computationally affordable, and the corresponding solution to the fixed-point equation is $\hat{v}_{TD}$. The other often used X is $X_{BR} = \Xi L \Phi$ [Scherrer, 2010], which we will not explained in details here. However, it is obvious that none of X_{TD} or X_{BR} is the optimal solution. In fact, it has been shown that $||V - \hat{v}_{TD}||_\Xi \leq \frac{1}{1-\gamma}||V - \hat{v}_{\text{best}}||_\Xi$ [Tsitsiklis and Van Roy, 1997], which implies that approximation error $||V - \hat{v}_{TD}||_\Xi$ between the true value function V and $\hat{v}_{TD}$ can be arbitrarily bad when $\gamma \to 1$. Given a fixed-point equation (15), we have

$$\Pi_\Phi^X T\hat{v} - \hat{v} = \Pi_\Phi^X (T\hat{v} - \hat{v}) = \Phi(X^T\Phi)^{-1}X^T(T\hat{v} - \hat{v}) \tag{16}$$

Thus we can formulate the approximation error via constraining $||X^T(T\hat{v} - \hat{v})||_\infty \leq \lambda$.

Since the P, R, Ξ models are not accessible, the sample-based estimation is used. Given n training samples of the form $\left(s_i, a_i, r_i, s_i'\right)_{i=1}^n$, $s_i \sim \Xi$, $r_i = r(s_i, a_i)$, $a_i \sim \pi_b(\cdot|s_i)$, $s_i' \sim P(\cdot|s_i, a_i)$, the empirical Bellman operator $\hat{T}$ is thus written as

$$\hat{T}(\hat{\Phi}\theta) = \hat{R} + \gamma\hat{\Phi}'\theta \tag{17}$$

We denote $\hat{\Phi}$ (resp. $\hat{\Phi}'$) the empirical feature matrices whose i-th row is the feature vector $\phi(s_i)^T$ (resp. $\phi(s_i')^T$), and $\hat{R} \in \mathbb{R}^n$ the empirical reward vector with corresponding r_i as the i-th row. And the error constraint can be formulated as $||Q^T(A\theta - b)||_\infty \leq \lambda$, where $A = \hat{\Phi} - \gamma\hat{\Phi}', b = \hat{R}$, and $Q \in \mathbb{R}^{n \times d}$. So given Q and a pre-defined error λ, the Dantzig Selector temporal difference learning problem can be formulated as

$$\hat{\theta}_{\text{DS-TD}} = \underset{\theta}{\text{argmin}} \|\theta\|_1 \quad \text{s.t.} \quad \|Q^T(A\theta - b)\|_\infty \leq \lambda.$$

If $Q = \hat{\Phi}$, then the algorithm is the DS-TD algorithm [Geist et al., 2012]. Motivated by the GDDS algorithm to find the optimal denoising matrix Q^T, we design a new Dantzig Selector temporal difference learning algorithm to compute the optimal Q^T for sparse value function approximation as follows

$$\hat{\theta}_{\text{ODDS-TD}} = \underset{\theta}{\text{argmin}} \|\theta\|_1 \quad \text{s.t.} \quad \|Q^T(A\theta - b)\|_\infty \leq \lambda \tag{18}$$

where Q is computed from solving

$$\min_Q : \ \|Q^T A - I\|_F^2 \quad \text{s.t.} : \quad \|Q_{.i}\| \leq 1 \tag{19}$$

Based on above, we propose the following ODDS-TD algorithm.

Algorithm 2 "Optimal" Denoising Dantzig Selector for Temporal Difference Learning with (ODDS-TD)

Require: A, b
Ensure: $\hat{\theta}_{\text{ODDS-TD}}$
Compute the denoising matrix Q^T via Eq. (19)
Compute $\hat{\theta}_{\text{ODDS-TD}}$ via Eq. (18)

5 Empirical Experiment

Experiments are first conducted on synthetic data with small-scale size and medium-scale respectively. Next, we apply the proposed method to a reinforcement learning problem on real data sets to show its advantage over existing algorithms.

5.1 Small-Scale Experiment

In this experiment, we conduct the comparison study between the regular Dantzig Selector (DS) and ODDS. We first compare the performance of different algorithms w.r.t different sparsity and noise levels. We set $n = 100, m = 150$, the true signal β^* is preset, with different numbers of non-zero elements (NNZ) among $10, 15, 20, 25$. We also change different noise level varying among $\sigma = 0.1, 0.2, 0.3$. Figure 3 shows the performance comparison of the two algorithms, wherein in each subfigure, the noise level is the same, and the x-axis denotes the NNZ level, and the result is measured by the difference of the learned sparse signal with β^*. From Figure 3, we can see that $||\hat{\beta}_{\text{ODDS}} - \beta^*||_2$ is much smaller than $||\hat{\beta}_{\text{DS}} - \beta^*||_2$.

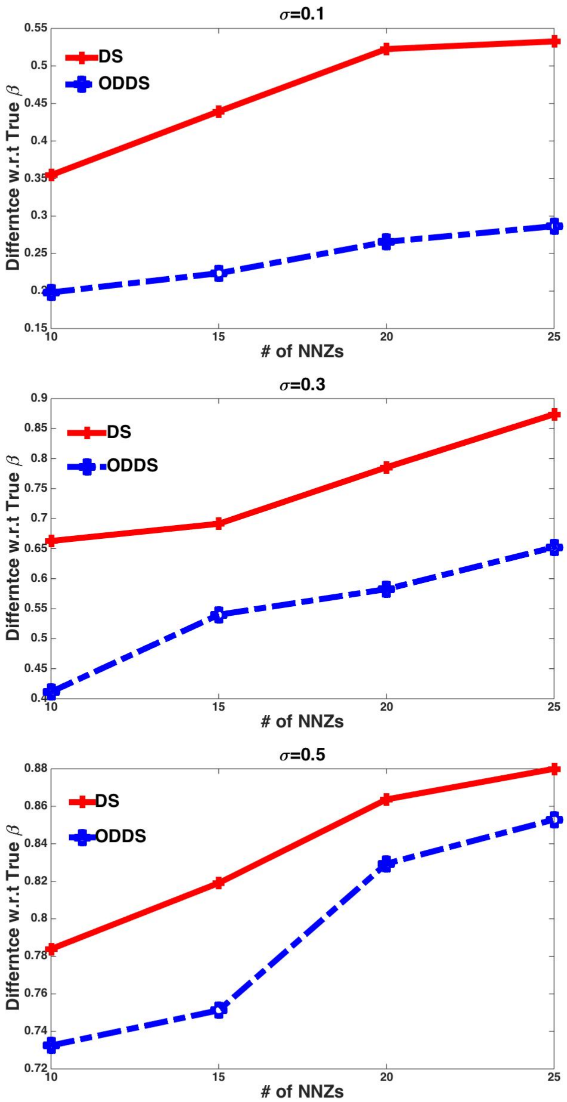

Figure 3: Small-scale Experiment w.r.t different sparsity levels.

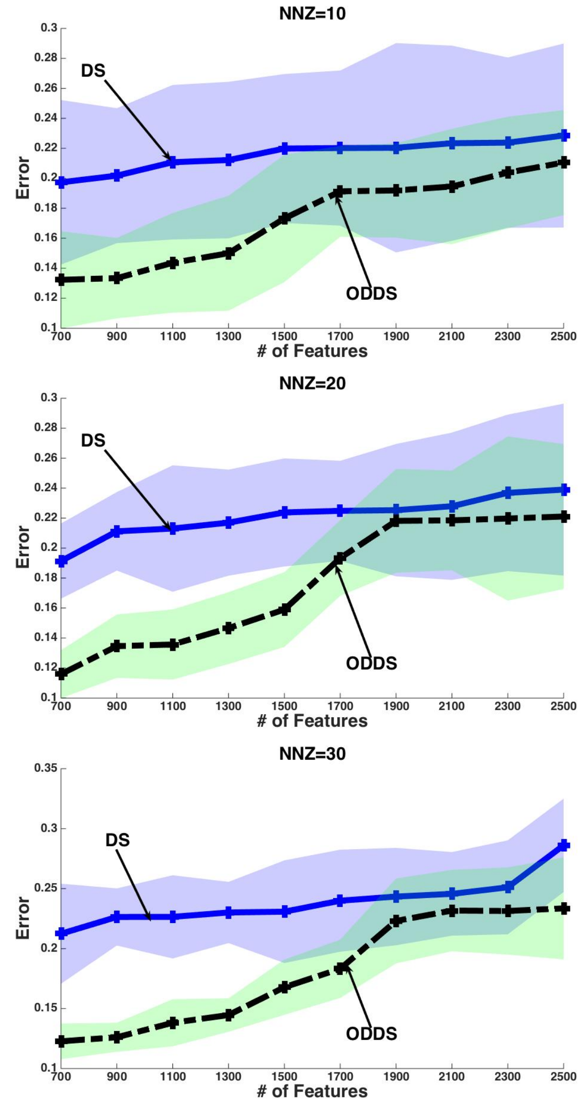

Figure 4: Comparison w.r.t different number of features.

5.2 Medium-Scale Experiment

In this experiment, we conduct the comparison study between the regular DS and ODDS. The experimental setting is that given $n = 500$, the number of features m goes among $700, 900, \cdots, 2500$, and $NNZ = 10$. The noise level $\sigma = 0.01$, and λ is chosen as $\lambda = \sigma\sqrt{2\log n}$ as suggested by [Candes and Tao, 2007] for a fair comparison between DS and ODDS, and the result is averaged by the mean-squares error (MSE), which is averaged over 50 runs. From Figure 4, we can see that the performance of ODDS is much better than that of regular DS, with both much lower MSE error and less variance.

5.3 ODDS-TD Experiment

In this experiment, we compare the performance of the DS-TD [Geist et al., 2012] and the proposed ODDS-TD algorithm. We test the performances of the two algorithms on the 20-state corrupted chain domain, wherein the two goal-states are $s \in [1, 20]$ with the reward signal $+1$, and the probability transition to the nearest goal-state is 0.9. The features are constructed as follows. 5 radial basis functions (RBF) are constructed, and one constant is used as an offset, and all other noisy features are randomly drawn from Gaussian distribution. So for each sample s_t, the feature vector $\phi(s) = [1, \mathrm{RBF}(1), \cdots, \mathrm{RBF}(5), s_1, \cdots, s_n]$. 200 off-policy samples are collected via randomly sampling the state space. Two comparison studies are carried out. In the first experiment, there are 300 noisy features, so altogether there are 306 features, and the value function approximation result is shown in the first subfigure of Figure 5, where v_{ODDS} and v_{DS} are the value function approximation results of ODDS and regular DS. From the figures, one can see that the v_{ODDS} is much more accurate than that of v_{DS}. The second experiment poses an even more challenging task, where the number of noisy features is set to be 600, so altogether there are 606 features, which makes the feature selection task much more difficult, and the result is shown in the second subfigure of Figure 5. We can see that

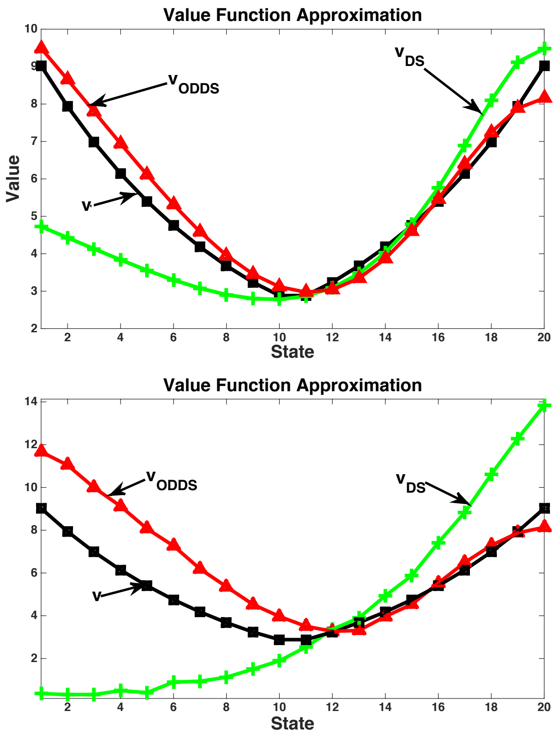

Figure 5: Comparison between ODDS-TD and DS-TD

v_{DS} has been severely distorted, whereas v_{ODDS} is still able to well preserve the topology of the value function, and is able to generate the right policy.

6 Conclusion

In this paper, motivated by achieving a better sparse signal recovery, we propose a generalized denoising matrix Dantzig selector formulation. A two-stage algorithm is proposed to find the optimal denoising matrix. The algorithm is then applied to sparse value function approximation problem in temporal difference learning field, and the empirical results validate the efficacy of the proposed algorithm. There are many interesting future directions along this research topic. For example, the ODDS framework can be extended to weighted Dantzig Selector (WDS) [Candes et al., 2008], multi-stage Dantzig Selector [Liu et al., 2010b], group Dantzig Selector [Liu et al., 2010a] and generalized Dantzig Selector [Chatterjee et al., 2014], which are parallel research directions of the Dantzig Selector algorithms family.

7 Acknowledgements

Bo Liu is partially funded by NSF Grant IIS-1216467, and the startup funding at Auburn University. Ji Liu is funded by NSF Grant CNS-1548078, the NEC fellowship, and the startup funding at the University of Rochester.

A Theoretical Analysis of ODDS

Proof to Theorem 1

Proof. Denote the difference between the solution to (5) $\hat{\beta}_{\text{GDDS}}$ and the true model β^* as $h = \hat{\beta}_{\text{GDDS}} - \beta^*$. Denote the support set of β^* by $T \subset \{1, 2, \cdots, m\}$. First we verify that the true model β^* is a feasible point to the problem (5) due to the following observation

$$\|Q^T(X\beta^* - y)\|_\infty = \|Q^T\epsilon\|_\infty \le \lambda.$$

Since $\hat{\beta}_{\text{GDDS}}$ is the minimizer of problem (5), it follows that

$$\begin{aligned}
&\|\hat{\beta}_{\text{GDDS}}\|_1 \le \|\beta^*\|_1 \\
\Rightarrow &\sum_{j\in T} |(\hat{\beta}_{\text{GDDS}})_j| + \sum_{j\in T^c} |(\hat{\beta}_{\text{GDDS}})_j| \le \sum_{j\in T} |\beta_j^*| \\
\Rightarrow &\sum_{j\in T^c} |(\hat{\beta}_{\text{GDDS}})_j| \le \sum_{j\in T} |\beta_j^* - (\hat{\beta}_{\text{GDDS}})_j| \\
\Rightarrow &\|h_{T^c}\|_1 \le \|h_T\|_1.
\end{aligned}$$

From the definition of $\rho(Q, X, p, s)$ in (6), we have

$$\rho(Q, X, p, \|\beta^*\|_0)\|h\|_p \le \|Q^T X h\|_\infty$$

which indicates

$$\|h\|_p \le \frac{\|Q^T X h\|_\infty}{\rho(Q, X, p, \|\beta^*\|_0)}. \tag{20}$$

It follows that

$$\begin{aligned}
\|Q^T X h\|_\infty =& \|Q^T X(\hat{\beta}_{\text{GDDS}} - \beta^*)\|_\infty \\
=& \|Q^T(X\hat{\beta}_{\text{GDDS}} - y + \epsilon)\|_\infty \\
\le& \|Q^T(X\hat{\beta}_{\text{GDDS}} - y)\|_\infty + \|Q^T\epsilon\|_\infty \\
\le& 2\|Q^T\epsilon\|_\infty.
\end{aligned}$$

Combining (20), we obtain the desired error bound

$$\|h\|_p \le \frac{2\|Q^T\epsilon\|_\infty}{\rho(Q, X, p, \|\beta^*\|_0)}.$$

It completes the proof. □

Proof of Lemma 1

Proof. This proof follows the standard union bound proof. For completion, we provide the proof below. Since $\epsilon \sim \mathcal{N}(0, I_{n\times n}\sigma^2)$, for each linear combination of ϵ, we have $Q_{.i}^T\epsilon \sim \mathcal{N}(0, \sigma^2)$. Recall that for any $t > 0$, we have that

$$\int_t^\infty e^{-x^2/2} dx \le e^{-t^2/2}/t. \tag{21}$$

Denote the event $A_i = \{|Q_{.i}^T\epsilon| \le \lambda\}, i = 1, 2, \cdots, m$:

$$\begin{aligned}
\{\|Q^T\epsilon\|_\infty \le \lambda\} &= \{\max_{1\le i\le m} |Q_{.i}^T\epsilon| \le \lambda\} \\
&= \cap_{i=1}^m A_i.
\end{aligned}$$

We derive the probability as follows:

$$\begin{aligned}\Pr(\cap_{i=1}^m A_i) &= 1-\Pr(\cup_{i=1}^m A_i^c)\\ &\geq 1-\sum_{i=1}^m \Pr(A_i^c)\\ &= 1-m\Pr(|\xi|\geq \lambda/\sigma)\\ &\geq 1-2\frac{m\sigma}{\lambda}e^{-\lambda^2/2\sigma^2},\end{aligned}$$

where the second inequality holds by union bound and the last inequality follows (21) since $\xi\sim\mathcal{N}(0,1)$. Thus, we can take $\lambda=2\sigma\sqrt{\log m}$ so that

$$\Pr(\|Q^T\epsilon)\|_\infty\leq\lambda)\geq 1-1/m\sqrt{\log m}=1-O(1/m).$$

It completes the proof. □

RIP Condition Implies GR Condition

Recall that to exactly recover β^* the RIP condition requires that $\delta_{2s}>\sqrt{2}-1$ [Candes, 2008], where $s=\|\beta^*\|_0$ and δ_{2s} is defined as for matrix X^TX is defined as

$$1-\delta_{2s}\leq\frac{\|Xg\|^2}{\|g\|^2}\leq 1+\delta_{2s}\quad\forall\,\|g\|_0\leq 2s.$$

Now we need to prove that $\delta_{2s}>\sqrt{2}-1$ leads to the GR condition $\rho(X,X,2,\|\beta^*\|_0)>0$.

Consider an arbitrary vector h and an arbitrary index set $T_0\subset\{1,2,\cdots,m\}$ with cardinality s satisfying $\|h_{T_0}\|_1\geq\|h_{T_0^c}\|_1$. T_1 corresponds to the locations of the s largest coefficients of $h_{T_0^c}$; T_2 to the locations of the next s largest coefficients of $h_{T_0^c}$, and so on. We use T_{01} to denote $T_0\cup T_1$ for short.

Note the fact that if $Xh\neq 0$, then $X^TXh\neq 0$. To obtain $\rho(X,X,2,\|\beta^*\|_0)>0$, it suffices to show that for any nonzero h satisfying $\|h_{T_0^c}\|_1\leq\|h_{T_0}\|_1$, the following ratio is positive

$$\|Xh\|/\|h\|>0. \tag{22}$$

We have

$$\begin{aligned}&|\langle Xh,Xh_{T_{01}}\rangle|\\ \geq&\|Xh_{T_{01}}\|^2-\sum_{j\geq 2}|\langle Xh_{T_{01}},Xh_{T_j}\rangle|\\ \geq&\|Xh_{T_{01}}\|^2-\sum_{j\geq 2}|\langle Xh_{T_0},Xh_{T_j}\rangle|-\sum_{j\geq 2}|\langle Xh_{T_1},Xh_{T_j}\rangle|\\ \geq&\|Xh_{T_{01}}\|^2-\delta_{2s}\sum_{j\geq 2}(\|h_{T_0}\|+\|h_{T_1}\|)\|h_{T_j}\|\\ \geq&\|Xh_{T_{01}}\|^2-\sqrt{2}\|h_{T_{01}}\|\delta_{2s}\sum_{j\geq 2}\|h_{T_j}\|\\ \geq&\|Xh_{T_{01}}\|^2-\sqrt{2}\|h_{T_{01}}\|\delta_{2s}\|h_{T_0^c}\|_1 s^{-1/2}\\ \geq&(1-\delta_{2s})\|h_{T_{01}}\|^2-\sqrt{2}\delta_{2s}\|h_{T_{01}}\|^2\\ \geq&(1-(\sqrt{2}+1)\delta_{2s})\|h_{T_{01}}\|^2,\end{aligned}$$

where the third inequality uses the result $|\langle Xh_{T_i},Xh_{T_j}\rangle|\leq\delta_{2s}\|h_{T_i}\|\|h_{T_j}\|$ if $i\neq j$, see Lemma 2.1 Candes [2008]). It follows from the fact $|\langle Xh,Xh_{T_{01}}\rangle|\leq\|Xh\|\|Xh_{T_{01}}\|$ that

$$\|Xh\|\geq(1-(\sqrt{2}+1)\delta_{2s})\|h_{T_{01}}\|. \tag{23}$$

From $\|h_{T_0^c}\|_1\leq\|h_{T_0}\|_1$, we have $\|h\|\leq 2\|h_{T_{01}}\|$, see the last line of the proof for Theorem 1.2 in [Candes, 2008]. It also can be found from [Candes and Tao, 2007, Bickel et al., 2009]. Together with (23) and the RIP condition $\delta_{2s}<\sqrt{2}-1$, we obtain

$$\|Xh\|\geq(1-(\sqrt{2}+1)\delta_{2s})\|h_{T_{01}}\|\geq\frac{(1-(\sqrt{2}+1)\delta_{2s})}{2}\|h\|,$$

which verifies (22).

Comparison to Existing Error Bounds for DS

This section aims to show that the error bound provided in (7) is a tighter bound than two existing results in [Candes and Tao, 2007] and [Bickel et al., 2009]. For simpler notations, we denote the difference between the estimate $\hat\beta_{\text{DS}}$ by DS and the true model β^* as $h=\hat\beta_{\text{DS}}-\beta^*$, T denotes the support set of β^*, and s denotes the sparsity of β^*. The complete comparison requires extensive space to basically repeat the proofs in [Candes and Tao, 2007] and [Bickel et al., 2009]. Here, we just show their results, highlight the key point in their original proofs, and illustrate why our error bound does not loose theirs.

Existing results conducted in [Candes and Tao, 2007] and [Bickel et al., 2009] are only based on the following facts

- (FACT 1) $\|h_{T^c}\|_1\leq\|h_T\|_1$ (Please refer to Eq. (3.2) in [Candes and Tao, 2007] and Eq. (B.12) in [Bickel et al., 2009]);
- (FACT 2) $\|X^TXh\|_\infty\leq 2\|X^T\epsilon\|_\infty\leq_{\text{(P)}}2\sigma\sqrt{\log m}$ (Please refer to Eq. (3.3) in [Candes and Tao, 2007] and Eq. (B.7) in [Bickel et al., 2009]).

which are also the foundations to provide our error bound in (7).

The error bound provided in Theorem 1.1 of [Candes and Tao, 2007] is derived from (see the end of Proof of Theorem 1.1 in [Candes and Tao, 2007])

$$\begin{aligned}\|h\|&\overset{\text{(FACT 1)}}{\leq}\frac{2\sqrt{s}}{1-\delta-\theta}\|X^TXh\|_\infty\\ &\overset{\text{(FACT 2)}}{\leq}\frac{2\sqrt{s}}{1-\delta-\theta}2\|X^T\epsilon\|_\infty.\end{aligned}\tag{24}$$

Please check the original paper for the definitions of δ and θ. The error bound (with $p=2$, $m=s$) provided in [Bickel et al., 2009] is derived from

$$\begin{aligned}\|h\|&\overset{\text{(FACT 1)}}{\leq}\frac{32s}{\kappa(s,s,1)}\|X^TX\epsilon\|_\infty\\ &\overset{\text{(FACT 2)}}{\leq}\frac{32s}{\kappa(s,s,1)}2\|X^T\epsilon\|_\infty.\end{aligned}\tag{25}$$

Please refer to the original paper for the definition of $\kappa(s, s, 1)$.

From these two bounds in (24) and (25), they essentially uses FACT 1 to find an upper bound for $\|h\|/\|X^T Xh\|_\infty$. This observation motivates us to directly define the upper bound through the definition of $\rho(\cdot, \cdot, \cdot)$ in (6)

$$\rho(X, X, s, 2) \leq \frac{\|X^T Xh\|_\infty}{\|h\|} \quad \forall h \text{ under FACT 1.}$$

Then we simply apply FACT 2 as (24) and (25) to obtain the error bound in (7). Therefore, our analysis provides a tighter error bound than Candes and Tao [2007] and Bickel et al. [2009].

References

S. R. Becker, E. J. Candès, and M. C. Grant. Templates for convex cone problems with applications to sparse signal recovery. *Mathematical Programming Computation*, 3 (3):165–218, 2011.

P. J. Bickel, Y. Ritov, and A. B. Tsybakov. Simultaneous analysis of lasso and dantzig selector. *The Annals of Statistics*, pages 1705–1732, 2009.

E. Candes and T. Tao. The Dantzig selector: Statistical estimation when p is much larger than n. *The Annals of Statistics*, pages 2313–2351, 2007.

E. J. Candes. The restricted isometry property and its implications for compressed sensing. *Comptes Rendus Mathematique*, 346(9):589–592, 2008.

E. J. Candes and T. Tao. Decoding by linear programming. *IEEE Transactions on Information Theory*, 51 (12):4203–4215, 2005.

E. J. Candes, M. B. Wakin, and S. P. Boyd. Enhancing sparsity by reweighted l1 minimization. *Journal of Fourier analysis and applications*, 14(5-6):877–905, 2008.

S. Chatterjee, S. Chen, and A. Banerjee. Generalized dantzig selector: Application to the k-support norm. In *Advances in Neural Information Processing Systems*, pages 1934–1942, 2014.

S. Chen and D. Donoho. Basis pursuit. In *IEEE the Twenty-Eighth Asilomar Conference onSignals, Systems and Computers*, volume 1, pages 41–44, 1994.

S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM review*, 43(1):129–159, 2001.

Y. Cong, J. Yuan, and J. Liu. Sparse reconstruction cost for abnormal event detection. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 3449–3456. IEEE, 2011.

D. L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006.

M. Geist, B. Scherrer, A. Lazaric, and M. Ghavamzadeh. A Dantzig Selector Approach to Temporal Difference Learning. In *International Conference on Machine Learning*, 2012.

B. Liu, S. Mahadevan, and J. Liu. Regularized off-policy TD-learning. In *Advances in Neural Information Processing Systems*, pages 845–853, 2012.

H. Liu, J. Zhang, X. Jiang, and J. Liu. The group dantzig selector. In *International Conference on Artificial Intelligence and Statistics*, pages 461–468, 2010a.

J. Liu. *Statistical Methods for Genome-wide Association Studies and Personalized Medicine*. PhD thesis, University of Wisconsin-Madison, 2014.

J. Liu, P. Wonka, and J. Ye. Multi-stage dantzig selector. In *Advances in Neural Information Processing Systems*, pages 1450–1458, 2010b.

J. Liu, R. Fujimaki, and J. Ye. Forward-backward greedy algorithms for general convex smooth functions over a cardinality constraint. *ICML*, 2013.

Z. Lu, T. K. Pong, and Y. Zhang. An alternating direction method for finding dantzig selectors. *Computational Statistics & Data Analysis*, 56(12):4037–4046, 2012.

S. Mahadevan and B. Liu. Sparse q-learning with mirror descent. *arXiv preprint arXiv:1210.4893*, 2012.

Y. Nesterov. *Introductory lectures on convex optimization: A basic course*, volume 87. Springer, 2004.

Z. Qin, W. Li, and F. Janoos. Sparse reinforcement learning via convex optimization. In *Proceedings of the 31st International Conference on Machine Learning (ICML-14)*, pages 424–432, 2014.

B. Scherrer. Should one compute the temporal difference fix point or minimize the bellman residual? the unified oblique projection view. In *ICML*, pages 52–68, 2010.

R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

J. A. Tropp. Greed is good: Algorithmic results for sparse approximation. *IEEE Transactions on Information Theory*, 50(10):2231–2242, 2004.

J. Tsitsiklis and B. Van Roy. An analysis of temporal-difference learning with function approximation. *IEEE Transactions on Automatic Control*, 42:674–690, 1997.

S. A. Van De Geer, P. Bühlmann, et al. On the conditions used to prove oracle results for the lasso. *Electronic Journal of Statistics*, 3:1360–1392, 2009.

X. Wang and X. Yuan. The linearized alternating direction method of multipliers for dantzig selector. *SIAM Journal on Scientific Computing*, 34(5):A2792–A2811, 2012.

H. Yu and D. P. Bertsekas. Error bounds for approximations from projected linear equations. *Mathematics of Operations Research*, 35(2):306–329, 2010.

T. Zhang. On the consistency of feature selection using greedy least squares regression. In *Journal of Machine Learning Research*, pages 555–568, 2009.

T. Zhang. Adaptive forward-backward greedy algorithm for learning sparse representations. *IEEE Transactions on Information Theory*, 57(7):4689–4708, 2011a.

T. Zhang. Sparse recovery with orthogonal matching pursuit under rip. *IEEE Transactions on Information Theory*, 57(9):6215–6221, 2011b.