Online Instrumental Variable Regression with Applications to Online Linear System Identification

Arun Venkatraman[1], Wen Sun[1], Martial Hebert[1], J. Andrew Bagnell[1], Byron Boots[2]
[1]Robotics Institute, Carnegie Mellon University, Pittsburgh, PA 15213
[2]School of Interactive Computing, Georgia Institute of Technology, Atlanta, GA 30332
{arunvenk,wensun,hebert,dbagnell}@cs.cmu.edu, bboots@cc.gatech.edu

Abstract

Instrumental variable regression (IVR) is a statistical technique utilized for recovering unbiased estimators when there are errors in the independent variables. Estimator bias in learned time series models can yield poor performance in applications such as long-term prediction and filtering where the recursive use of the model results in the accumulation of propagated error. However, prior work addressed the IVR objective in the batch setting, where it is necessary to store the entire dataset in memory - an infeasible requirement in large dataset scenarios. In this work, we develop Online Instrumental Variable Regression (OIVR), an algorithm that is capable of updating the learned estimator with streaming data. We show that the online adaptation of IVR enjoys a no-regret performance guarantee with respect the original batch setting by taking advantage of any no-regret online learning algorithm inside OIVR for the underlying update steps. We experimentally demonstrate the efficacy of our algorithm in combination with popular no-regret online algorithms for the task of learning predictive dynamical system models and on a prototypical econometrics instrumental variable regression problem.

Introduction

Instrumental variable regression (IVR) is a popular statistical linear regression technique to help remove bias in the prediction of targets when both the features and targets are correlated with some unknown additive noise, usually a variable omitted from the regression due to the difficulty in observing it (Bowden and Turkington 1990). In this setting, ordinary least squares (OLS) (i.e. linear regression) from features to targets leads to a biased estimate of the dependence between features and targets. For applications where the underlying unbiased dependency is required, such as in the study of causal effects for econometrics (Miguel, Satyanath, and Sergenti 2004), epidemiology (Greenland 2000), or for the learning of dynamical system models (Söderström and Stoica 2002), IVR provides a technique to remove the correlation with the unobserved variables.

We focus in this work on the regression application of instrumental variables where the IVR process consists of multiple linear regressions steps. Prior attention on IVR has focused on the batch learning scenario: each step of regression is performed in whole with all of the data at once. However, with the ever growing prevalence of large datasets, such an approach becomes quickly infeasible due to the scaling of the memory and computational complexity with regards to the data set size and feature dimensionality. Towards this end, we propose an online version of instrumental variable regression that replaces each of the regression steps with an online learner.

Specifically, we develop an Online Instrumental Variable Regression (OIVR) procedure that can be regarded as a reduction to no-regret online learning. Under the assumption that the set of regression and instrumental variables are i.i.d, we derive a strong no-regret bound with respect to the desired objective optimized by the batch setting (batch IVR). Our theorem allows us to take advantage of *any* no-regret online learning procedure for the multiple regression steps in IVR. We explicitly show that OIVR allows us to introduce a new family of online system identification algorithms that can exploit no-regret online learning. This reduction extends on the initial reduction given by Hefny et. al (Hefny, Downey, and J. Gordon 2015) from batch predictive state dynamical system learning to batch IVR. Finally, we investigate the experimental performance of several popular online algorithms such as Online Gradient Descent (OGD) (Zinkevich 2003), Online Newton Step (Hazan, Agarwal, and Kale 2006) (ONS), Implicit Online Gradient Descent (iOGD) (Kulis et al. 2010), and Follow The Regularized Leader (FTRL) (Shalev-Shwartz 2011) in the context of OIVR for both dynamical system modeling and on a simple but illustrative econometrics example.

Instrumental Variable Regression

Consider the standard linear regression scenario where we wish to find A given design matrices (datasets) $X = [x_1 \quad x_2 \quad \ldots]$ and $Y = [y_1 \quad y_2 \quad \ldots]$ representing our explanatory variables (features) $x_i \in \mathbb{R}^{n\times 1}$ and outputs (targets) $y_i \in \mathbb{R}^{m\times 1}$. This relationship is modeled by:

$$Y = AX + E \tag{1}$$

where $E = [\varepsilon_1 \quad \varepsilon_2 \quad \ldots]$ are independent noise (error).

Solving this via least-squares minimization, gives us:

$$\begin{aligned}\hat{A} &= YX^T(XX^T)^{-1} = (AX+E)X^T(XX^T)^{-1} \\ &= AXX^T(XX^T)^{-1} + EX^T(XX^T)^{-1} \\ &= A + EX^T(XX^T)^{-1} \end{aligned} \tag{2}$$

When the number of samples T goes to infinity, by law of large number, we will have: $EX^T/T \rightarrow \mathbf{E}(\varepsilon x^T)$ and $XX^T/T \rightarrow \mathbf{E}(xx^T)$ in probability. Normally, we assume that ε and x are uncorrelated, which means EX^T/T converges to zero in probability, ($\mathbf{E}(\varepsilon x^T) = 0$), which yields an unbiased estimate of A from Eq. 2.

$$\hat{A} = A + \frac{1}{T}EX^T\left(\frac{1}{T}XX^T\right)^{-1} \rightarrow A$$

However, if ε and x are correlated $\mathbf{E}(\varepsilon x^T) \neq 0$, we are only able to get a biased estimate of A through the least-squares optimization, since $\mathbf{E}\left[\varepsilon x^T\right]\mathbf{E}\left[xx^T\right]^{-1} \neq 0$.

On the other hand, IVR can achieve an unbiased estimate of A (Rao et al. 2008; Cameron and Trivedi 2005). In IVR, we remove this bias by utilizing an *instrumental variable*, denoted as $Z = [z_1 \quad z_2 \quad \ldots]$ in Fig. 1. For a variable to be an instrumental variable, we need two conditions: (1) the instrumental variable z is correlated with x such that $\mathbf{E}(xz^T)$ is full row rank and (2) the instrumental variable is uncorrelated with ε, i.e. $\mathbf{E}(z\varepsilon^T) = 0$.

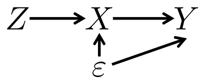


Figure 1: Causal diagram for IVR

Instrumental variable regression proceeds as follows. IVR first linearly regresses from Z to X to get $\hat{X} = XZ^{\dagger}Z$, where $Z^{\dagger} = Z^T(ZZ^T)^{-1}$. Then, IVR linearly regresses from the projection $\hat{X}$ to Y to get an estimate of A:

$$\begin{aligned}\hat{A}_{\text{IVR}} &= Y\hat{X}^T(\hat{X}\hat{X}^T)^{-1} \\ &= YZ^{\dagger}ZX^T(XZ^{\dagger}ZX^T)^{-1} \\ &= AXZ^{\dagger}ZX^T(XZ^{\dagger}ZX^T)^{-1} \\ &\quad + EZ^{\dagger}ZX^T(XZ^{\dagger}ZX^T)^{-1} \\ &= A + EZ^T(ZZ^T)^{-1}ZX^T(XZ^{\dagger}ZX^T)^{-1}\end{aligned}$$

Note that $\hat{A}_{IVR} \rightarrow A$ in probability since $EZ^T/T \rightarrow 0$ in probability under the assumption that instrumental variable z and ε are uncorrelated.

The process of instrumental variable regression can be represented through the following two steps (also known as two-stage regression):

$$M^* \leftarrow \underset{M}{\arg\min} \|X - MZ\|_F^2 \tag{3}$$

$$A^* \leftarrow \underset{A}{\arg\min} \|Y - AM^*Z\|_F^2 \tag{4}$$

For shorthand, we will generally refer to the final regression stage, Eqn. 4, as the batch IVR objective .

Algorithm 1 Batch Instrumental Variable Regression

```
Input:
    ▷ Explanatory Variable Design Matrix $X \in \mathbb{R}^{d_x,n}$,
    ▷ Instrumental Variable Design Matrix $Z \in \mathbb{R}^{d_z,n}$,
    ▷ Prediction Targets Design Matrix $Y \in \mathbb{R}^{d_y,n}$
Output: $A^* \in \mathbb{R}^{d_y,d_x}$
    $M^* \leftarrow \arg\min_M \|X - MZ\|_F^2$
    $\widehat{X} \leftarrow M^*Z$
    $A^* \leftarrow \arg\min_A \left\|Y - A\widehat{X}\right\|_F^2$
    return $A^*$
```

Online Instrumental Variable Regression

The formulation of online algorithms yields a two-fold benefit – first, it allows us to use datasets that are too large to fit in memory by considering only one or a few data points at a time; second, it allows us to run our algorithm with streaming data, a vital capability in fields such as robotics where many sensors can push out volumes of data in seconds. In the following section, we formulate an online, streaming-capable adaptation of the batch Instrumental Variable Regression (IVR) algorithm. We show that our online algorithm has a strong theoretical performance guarantee with respect to the performance measure in the batch setting.

In the batch setup of IVR (Algorithm 1), we require all the datapoints *a priori* in order to find A^*. To create an online version of this algorithm, it must instead compute estimates M_t and A_t as it receives a single set of data points, x_t, z_t, y_t. To motivate our adaptation of IVR, we first consider the adaptation of OLS (i.e. linear regression) to the online setting. Given the design matrices $X = [x_0, \ldots, x_t, \ldots,]$ and $Y = [y_0, \ldots, y_t, \ldots]$, in OLS, we optimize the following batch objective over all the data points:

$$\beta^* = \underset{\beta}{\arg\min} \|\beta X - Y\|_F^2 = \underset{\beta}{\arg\min} \sum_t \ell_t(\beta) \tag{5}$$

where the loss function $\ell_t(\beta) = \|\beta x_t - y_t\|_2^2$ is the L2 loss for the corresponding pair of data points (x_t, y_t). To formulate an online OLS algorithm, we may naturally try to optimize the L2 loss for an individual data point pair $\|\beta x_t - y_t\|_2^2$ at each timestep without directly considering the loss induced by other pairs. Prior work in the literature has developed algorithms that address this problem of considering losses ℓ_t and predicting a β_{t+1} while still achieving provable performance with respect to the optimization of β over the batch objective (Zinkevich 2003; Hazan, Agarwal, and Kale 2006; Shalev-Shwartz 2011). The performance guarantee of these algorithms is in terms of the (average) regret, which is defined as:

$$\frac{1}{T}\text{REGRET} = \frac{1}{T}\sum_t \ell_t(\beta_t) - \frac{1}{T}\min_{\beta}\sum_t \ell_t(\beta) \tag{6}$$

We say a learning procedure is *no-regret* if $\lim_{T\rightarrow\infty} \frac{1}{T}(\text{REGRET}) = 0 \Rightarrow \text{REGRET} \in o(T)$.

Intuitively, the no-regret property tells us that the optimization of the the loss in this way gives us a solution that is competitive with the best result in hindsight (i.e. if we had optimized over the losses from all data points). In IVR (Algorithm 1), lines 1 and 3 are each linear regression steps which are individually the same as Eqn. 5. Motivated by this, we introduce Online Instrumental Variable Regression (OIVR), in which we utilize a no-regret online learner for the individual batch linear regressions in IVR. The detailed flow of OIVR is shown in Algorithm 2.

Algorithm 2 Online Instrumental Variable Regression with No-Regret Learners

```
Input:
    ▷ no-regret online learning procedures LEARN$_1$, LEARN$_2$
    ▷ Streaming data sources for the explanatory variable $S_x(t) : t \to x \in \mathbb{R}^{d_x}$, the instrumental variable $S_z(t) : t \to z \in \mathbb{R}^{d_z}$, and the target variable $S_y(t) :\to y \in \mathbb{R}^{d_y}$
Output: $\bar{A}_T \in \mathbb{R}^{d_y,d_x}$
 1: Initialize $M_0 \in \mathbb{R}^{d_x,d_z}$, $A_0 \in \mathbb{R}^{d_y,d_x}$
 2: Initialize $\bar{M}_0 \leftarrow \mathbf{0} \in \mathbb{R}^{d_x,d_z}$, $\bar{A}_0 \leftarrow \mathbf{0} \in \mathbb{R}^{d_y,d_x}$
 3: Initialize $t \leftarrow 1$
 4: while $S_x \neq \emptyset$ and $S_z \neq \emptyset$ and $S_y \neq \emptyset$ do
 5:     $(x_t, z_t, y_t) \leftarrow (S_x(t), S_z(t), S_y(t))$
 6:     $M_t \leftarrow$ LEARN$_1(z_t, x_t, M_{t-1})$
 7:     $\bar{M}_t \leftarrow ((t-1)\bar{M}_{t-1} + M_t)/t$
 8:     $\widehat{x}_t \leftarrow \bar{M}_t z_t$
 9:     $A_t \leftarrow$ LEARN$_2(\widehat{x}_t, \widehat{y}_t, A_{t-1})$
10:     $\bar{A}_t \leftarrow ((t-1)\bar{A}_{t-1} + A_t)/t$
11:     $t \leftarrow t + 1$
12: end while
13: return $\bar{A}_t$
```

From the definition of no-regret for the optimization on lines 6 and 9 in Algorithm 2, we get the following:

$$\frac{1}{T}\sum_t \|M_t z_t - x_t\|_F^2 - \frac{1}{T}\min_M \sum_t \|M z_t - x_t\|_F^2 \le o(T)$$

$$\frac{1}{T}\sum_t \left\|A_t \bar{M}_t z_t - y_t\right\|_F^2 - \frac{1}{T}\min_A \sum_t \left\|A \bar{M}_t z_t - y_t\right\|_F^2 \le o(T)$$

Though these regret bounds give us a guarantee on each individual regression with respect the sequence of data points, they fail to give us the desired performance bound as we get in the single OLS scenario; these bounds do not show that this method is competitive with the optimal result from batch IVR in hindsight (e.g., how close is $\bar{A}_t$ to A^* from Algorithm 1 with M^* instead of $\bar{M}_t$). We wish to show that this algorithm is in fact competitive with the batch instrumental variable regression algorithm (Algorithm 1). Specifically, we focus on the stochastic setting where each set of data points $(x_t, z_t, y_t) \sim P$ is i.i.d.. In this setting, we would like to show that:

$$\sum_t \mathbf{E}\left[\left\|\bar{A}_t M^* z - y\right\|_2^2\right] - \min_A \sum_t \mathbf{E}\left[\|AM^* z - y\|_2^2\right] \le o(T)$$

where $\min_A \sum_t \mathbf{E}\left[\|AM^*z - y\|_2^2\right]$ is exactly the last objective of batch IVR, since under the assumption that x_t, y_t, z_t are i.i.d, $\frac{1}{T}\|Y - AM^*Z\|_F^2$ (Eq. 4) converges to $\mathbf{E}\left[\|AM^*z - y\|_2^2\right]$ in probability. In the below section, we derive the above bound for our OIVR algorithm. Through this bound, we are able to show that even though we optimize A_t with regards to the *online* $\bar{M}_t$ at every timestep, we are finally competitive with the solution achieved by the batch procedure that uses the *batch* M^* to learn A^*. We also note that the prior technique, recursive IVR (e.g. (Söderström and Stoica 2002)), is similar to using FTRL (Shalev-Shwartz 2011) with rank-1 updates. We below extend prior analysis in that we derive a regret bound for this type of update procedure.

Performance Analysis of Online IVR

In order to derive the primary theoretical contribution of this work in Theorem 2, we first present the lemma below with a derivation in the appendix (supplementary material). We follow with a sketch of the proof for the performance guarantee of OIVR with regards to the batch IVR solution and recommend the reader to the appendix for the detailed derivation.

In lines 7 and 10 of Algorithm 2, we compute an average of the sequence of predictors (matrices). This computation can be done relatively efficiently without storing all the predictors trained. The usefulness of this operation can be seen in the result of the below lemma.

Lemma 1. *Given a sequence of convex loss functions $\{\ell_t(\beta)\}$, $1 \le t \le T$, and the sequence of $\{\beta_t\}$ that is generated by any no-regret online algorithm, under the assumption that ℓ_t is i.i.d and $\ell = \mathbf{E}(\ell_t)$, the average $\bar{\beta} = 1/T \sum_t \beta_t$ of $\{\beta_t\}$ has the following properties:*

$$\mathbf{E}\left[\ell(\bar{\beta}) - \ell(\beta^*)\right] \le \frac{1}{T} r(T) \to 0, \;\; T \to \infty, \tag{7}$$

where $r(T)$ stands for the function of the regret bound with respect to T[1], which is sublinear and belongs to $o(T)$ for all no-regret online learning algorithms. When ℓ is α strongly convex with respect to β in norm $\|\cdot\|$, we have:

$$\mathbf{E}\left[\|\bar{\beta} - \beta^*\|\right] \le \frac{2}{\alpha T} r(T) \to 0, \;\; T \to \infty \tag{8}$$

Similar online-to-batch analysis can be found in (Cesa-Bianchi, Conconi, and Gentile 2004; Littlestone 2014; Hazan and Kale 2014). For completeness, we include the proof of the lemma in the appendix.

With this, we now approach the main theorem for the regret bound on Online Instrumental Variable Regression. We explicitly assume that x_t, y_t and z_t are i.i.d, and $x = \mathbf{E}(x_t)$, $y = \mathbf{E}(y_t)$, $z = \mathbf{E}(z_t)$, and $\mathbf{E}(z_t z_t^T)$ is positive definite.

Theorem 2. *Assume (x_t, y_t, z_t) are i.i.d. and $\mathbf{E}(zz^T)$ is postive definite. Following any online no-regret procedure on the convex L2 losses for M_t, and A_t and computing $\bar{M}_t$, $\bar{A}_t$ as shown in Algorithm 2, we get that as $T \to \infty$:*

$$\mathbf{E}\left[\left\|\bar{A}_T M^* z - y\right\|_2^2\right] \to \mathbf{E}\left[\left\|A^* M^* z - y\right\|_2^2\right] \tag{9}$$

$$\text{and} \quad \bar{A}_T \to A^* \tag{10}$$

for the A^*, M^* from Batch IVR *(Alg. 1).*

[1] For instance, online gradient descent (Zinkevich 2003) has $r_t(T) = C\sqrt{T}$ for some positive constant C.

Proof. For the sake of brevity, we provide the complete proof in the appendix and an abbreviated sketch below.

Since we run a no-regret online algorithm for A_t on loss function $\|A_t\bar{M}_t z_t - y_t\|_2^2$, we have:

$$\sum_t \|A_t\bar{M}_t z_t - y_t\|_2^2 \le \textsc{Regret}_A + \sum_t \|A^*\bar{M}_t z_t - y_t\|_2^2$$

where $\sum_t$ denotes $\sum_{t=1}^T$.

Let $\epsilon_t = M^* - \bar{M}_t$. Then, expanding the squared norms on the left and right side of the inequality, rearranging terms, and upperbounding the terms we get:

$$\begin{aligned}
\sum_t \|A_t M^* z_t - y_t\|_2^2 &\le \textsc{Regret}_A \\
&+ \sum_t \|A^* M^* z_t - y_t\|_2^2 \\
&+ \|A^*\|_F^2 \|\epsilon_t\|_F^2 \|z_t\|_2^2 + \|A_t\|_F^2 \|\epsilon_t\|_F^2 \|z_t\|_2^2 \\
&+ 2|(A^* M^* z_t - y_t)^T (A^*\epsilon_t z_t)| \\
&+ 2|(A_t M^* z_t - y_t)^T (A_t\epsilon_t z_t)|
\end{aligned}$$

Assume that $\|z_t\|_2$, $\|y_t\|_2$, $\|M^*\|_F$, $\|A_t\|_F$, $\|A^*\|_F$ are each always upper bounded by some positive constant. Defining positive constants C_1 and C_2 appropriately and using the Cauchy-Swartz and triangle inequalities, we get:

$$\begin{aligned}
\sum_t \|A_t M^* z_t - y_t\|_2^2 &\le \textsc{Regret}_A \qquad (11) \\
+ \sum_t \|A^* M^* z_t - y_t\|_2^2 &+ C_1 \|\epsilon_t\|_F^2 + C_2 \|\epsilon_t\|_F
\end{aligned}$$

Since we run a no-regret online algorithm on loss $\|M_t z_t - x_t\|_2^2$ with the assumptions that z_t, x_t, and y_t are i.i.d and $\mathbf{E}[zz^T]$ is positive definite, we get as $t \to \infty$:

$$\mathbf{E}\,\|\epsilon_t\|_F^2 \le \frac{1}{t} r_M(t) \to 0 \text{ and } \mathbf{E}\,\|\epsilon_t\|_F \le \sqrt{\frac{1}{t} r_M(t)} \to 0,$$

where $\mathbf{E}$ is the expectation under the randomness of the sequences z and x. Considering the stochastic setting (i.e. i.i.d z_t, x_t, and y_t), applying Cesaro Mean (Hardy 2000) and taking $T \to \infty$:

$$\frac{1}{T}\mathbf{E}\left[\sum_t \|A_t M^* z - y\|_2^2\right] \le \mathbf{E}\left[\|A^* M^* z - y\|_2^2\right]$$

Thus, we have shown the algorithm is no-regret.

Let $\bar{A}_T = \frac{1}{T}\sum_t A_t$. Using Jensen's inequality, we get:

$$\mathbf{E}\left[\|\bar{A}_T M^* z - y\|_2^2\right] \le \mathbf{E}\left[\|A^* M^* z - y\|_2^2\right]$$

Since the above is valid for any A^*, let $A^* = \arg\min_A \mathbf{E}\left[\|AM^* z - y\|_2^2\right]$. Due to bounding from above and below by the objective at A^*, we get:

$$\mathbf{E}\left[\|\bar{A}_T M^* z - y\|_2^2\right] \to \mathbf{E}\left[\|A^* M^* z - y\|_2^2\right]$$

With $\mathbf{E}\left[zz^T\right] \succ 0$ resulting in strongly convex objective, we get a unique minimizer for the objective:

$$\bar{A}_T \to A^*, \;\; T \to \infty$$

□

We also want to note that the regret rate of our algorithm depends on the no-regret online algorithms used. For instance, if we use OGD, which has no-regret rate of $O(\sqrt{T}/T)$ for $\textsc{Learn}_1$ and $\textsc{Learn}_2$, then our algorithm has a no-regret rate of $O(\sqrt{T}/T)$. The choice of learning algorithm is related to the desired trade-off between computational complexity and convergence rate. FTRL and ONS can have faster convergence, making them suitable for applications where obtaining samples is difficult: e.g., data from a physical robot. In contrast, gradient-based algorithms (e.g. iOGD, OGD) have lower computational complexity but may converge slower, making them useful for scenarios where obtaining samples is cheap, e.g., data from video games.

Dynamical Systems as Instrumental Variable Models

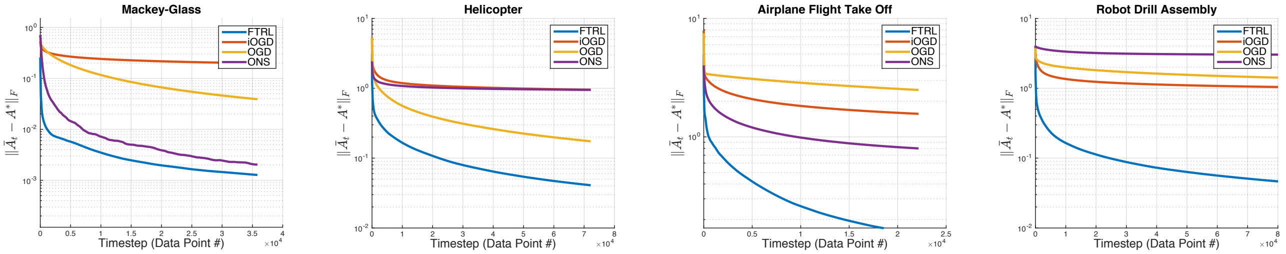
For a dynamical system, let us define state $s \in \mathcal{S} \in \mathbb{R}^m$ and observation $o \in \mathcal{O} \in \mathbb{R}^n$. At time step t, the system stochastically transitions from state s_t to state s_{t+1} and then receives an observation o_{t+1} corresponding to s_{t+1}. A dynamical system generates a sequence of observations o_t from latent states s_t connected in a chain. A popular family of algorithms for representing and learning dynamical systems are predictive state representations (PSRs) (Littman, Sutton, and Singh 2001; Singh, James, and Rudary 2004; Boots and Gordon 2012; 2011a; 2011b; Boots, Siddiqi, and Gordon 2011; Hefny, Downey, and J. Gordon 2015). It also has been shown in (Boots and Gordon 2011b; Hefny, Downey, and J. Gordon 2015) that we can interpret the problem of learning PSRs as linear instrumental-variable regression, which reduces the dynamical system learning problem to a regression problem.

Following (Hefny, Downey, and J. Gordon 2015), we define the predictive state Q as $Q_t = \mathbf{E}(o_{t:t+k-1}|o_{1:t-1})$ (instead of tracking the posterior distribution $\mathbf{P}(s_t|o_{1:t-1})$ on state, we track the observable representation Q_t), where $o_{t:t+k-1}$ is a k-step time window of *future* observations. We also define the *extended future* observations as $o_{t:t+k}$, which is a $(k+1)$-step time window of future observations. The predictive state representation of extended futures is defined as $P_t = \mathbf{E}(o_{t:t+k}|o_{1:t-1})$. Therefore, learning a dynamical system is equivalent to finding an operator A that maps from Q_t to P_t:

$$P_t = AQ_t \qquad (12)$$

With A and the initial belief $Q_0 = \mathbf{E}(o_{0:k-1})$, we are able to perform filtering and prediction. Given the belief Q_t at step t, we use A to compute $P_t = \mathbf{E}(o_{t:t+k}|o_{1:t-1})$. To compute $\mathbf{E}(o_{t+1:t+k}|o_{1:t-1})$ (prediction), we simply drop the o_t from P_t. For filtering, given a new observation o_t, under the assumption that the extended future $o_{t:t+k}$ has constant covariance, we can compute $\mathbf{E}(o_{t+1:t+k}|o_{1:t})$ by simply performing a conditional Gaussian operation.

A naive approach to compute A is to use ordinary linear regression directly from futures $o_{t:t+k-1}$ to extended futures $o_{t:t+k}$. However, even though $o_{t:t+k-1}$ and $o_{t:t+k}$ are unbiased samples of Q_t and P_t, they are noisy observations of Q_t and P_t respectively. The noises overlap: $o_{t:t+k-1}$ and $o_{t:t+k}$ share a k-step time window (Hefny, Downey, and J. Gordon 2015). Therefore, directly regressing from Q_t to P_t gives a



(a) Mackey-Glass ($\tau = 10$)
(b) Helicopter
(c) Airplane Flight Take Off
(d) Robot Drill Assembly

Figure 2: Convergence Plots for the Dynamical System Experiments. (Best viewed in color)

biased estimate of A, which can lead to poor prediction and filtering performance. Indeed, as verified in our experiments, the biased A computed by ordinary least square regression performs worse in comparison to the IVR based methods.

To overcome the bias, the authors in (Boots and Gordon 2011b; Hefny, Downey, and J. Gordon 2015) introduce past observations $o_{t-k:t-1}$ as instruments. The past observations $o_{t-k:t-1}$ are not correlated with the *noise* in the future observations $o_{t:t+k-1}$ and extended future observations $o_{t:t+k}$ but are correlated with Q_t. Explicitly matching terms to those used in IVR, the noisy observation of P_t is equivalent to y, the noisy observation of Q_t is equivalent to x, and the past observations $o_{t-k:t-1}$ is the instrumental variable z. Unlike (Hefny, Downey, and J. Gordon 2015), which introduced a batch IVR (Alg. 1) for system identification, we focus on using OIVR (Alg. 2) where we receive observations online.

Online Learning for Dynamical Systems

Given OIVR, learning dynamical systems online becomes straightforward. To apply Alg. 2 to model dynamical systems, we maintain a k-step time window of the *future* $o_{t:t+k-1}$, a $(k+1)$-step time window of the *extended future* $o_{t:t+k}$, and a k-step time window of the *past* $o_{t-k:t-1}$. Matching terms to Alg. 2, we set $x_t = o_{t:t+k-1}$, $y_t = o_{t:t+k}$, and $z_t = o_{t-k:t-1}$. With x_t, y_t and z_t, we update M_t and A_t following lines 6 and 9. When a new observation o_{t+k+1} is received, the update of x_t and y_t and z_t to x_{t+1}, y_{t+1} and z_{t+1} is simple and can be computed efficiently (e.g., to compute $y_{t+1} = o_{t+1:t+k+1}$, we simply drop o_t from x_t and append o_{t+k+1} at the end (i.e. circular buffer)).

By maintaining these three fixed-step time window of observations instead of building a large Hankel matrix ((Hefny, Downey, and J. Gordon 2015; Boots, Siddiqi, and Gordon 2011)) that stores concatenations of all the observations, we significantly reduce the required space complexity. At every online update step (lines 6 and 9 in Alg. 2), the online learning procedure usually has lower computational complexity. For instance, using Online Gradient Descent (Zinkevich 2003) makes each step requires $O((kn)^2)$ compared to the $O((kn)^3)$ in the batch-based algorithms (usually due to matrix inversions).

Experiments

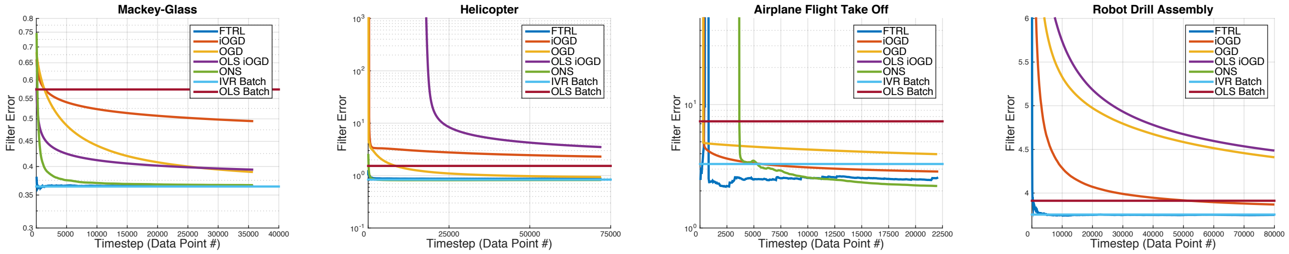
We demonstrate the performance of OIVR on a variety of dynamics benchmark and one illustrative econometrics problem. In Fig. 2, we show the convergence of the estimated $\bar{A}_t$ in OIVR to the A^* computed with IVR. As an additional performance metric, we report the observation prediction error with a constant covariance Kalman filter using $\bar{A}_t$ (Fig. 3) on a set of held out test trajectories. For computational reasons, we report the filter error after every 50 data points given to the online learner. Below we describe each our test benches.

MG-10 The Mackey-Glass (MG) time-series is a standard dynamical modelling benchmark (Ralaivola and D'Alche-Buc 2004; Wingate and Singh 2006) generated from the nonlinear time-delay differential equation $\dot{x}(t) = -bx(t) + \frac{ax(t-\tau)}{1+x(t-\tau)^{10}}$. This system produces chaotic behavior for larger time delays τ (seconds). We generated 30 trajectories with random initializations and $\tau = 10$, $a = 0.7$, $b = 0.35$.

Helicopter The simulated helicopter from (Abbeel and Ng 2005) computes its dynamics in a 21-dimensional state space with a 4-dimensional control input. In our experiments, a closed loop LQR controller attempts to bring the helicopter to hover at a fixed point from randomly chosen starting configurations. White noise is added in each state transition. The LQR controller chooses actions based on state and it poses a challenge for the learner to extract this implicit relationship governing the evolution of the system.

Airplane Flight Take Off We also consider the complex dynamics generated during a DA-42 airplane's take off in a flight simulator, X-plane (Research 2015), a well known program for training pilots. Trajectories of observations, which include among others speed, height, angles, and the pilot's control inputs, were collected were collected from a human expert controlling the aircraft. Due to high correlation among the observation dimensions, we precompute a whitening projection at the beginning of online learning using a small set of observations to reduce the dimensionality of the observations by an order of magnitude.

Robot Drill Assembly Our final dynamics benchmark consists of 96 sensor telemetry traces from a robotic



(a) Mackey-Glass ($\tau = 10$)

(b) Helicopter

(c) Airplane Flight Take Off

(d) Robot Drill Assembly

Figure 3: Filtering Error for the Dynamical System Experiments. Note that the results for OLS iOGD in Fig. 3(c) and for ONS in Fig. 3(d) are higher than the plotted area. (Best viewed in color)

manipulator assembling the battery pack on a power drill. The 13 dimensional observations consist of the robot arm's 7 joint torques as well as the 3D force and torque vectors as measured at the wrist of the robotic arm. The difficulty in this real-world dataset is that the sensors, especially the force-torque sensor, are known to be noisy and are prone to hysteresis. Additionally, the fixed higher level closed-loop control policy for the drill assembly task is not explicitly given in the observations and must be implicitly learned.

We applied four different no-regret online learning algorithms on the dynamical system test benches: OGD (Online Gradient Descent), iOGD (implict Online Gradient Descent), ONS (Online Newton Step), and FTRL (Follow the Regularized Leader). Fig. 2 shows the convergence of these online algorithms in terms of $\|\bar{A}_t - A^*\|_F$, where $\bar{A}_t$ is the solution of OIVR at time step t and A^* is the solution from batch IVR. Though FTRL had the fastest convergence, FRTL is memory intensive and computationally expensive as it runs a batch IVR at every time step over all the data points which have to be stored. ONS, also a computationally intensive algorithm, generally achieved fast convergence on the testbenches, except in the Robot Drill Assembly benchmark due to the difficulty in tuning the parameters of ONS. In general, OGD and iOGD perform well while only requiring storage of the latest data point. Furthermore, these algorithms have lower computational complexity than ONS and FTRL at each iteration.

We also compared the filtering performance of these OIVR methods with batch IVR, batch OLS, and online OLS (via iOGD) on these datasets. The results are shown in Fig. 3. First, by comparing the batch IVR and batch OLS, we observe that the biased A computed by batch OLS is consistently outperformed on the filtering error by the A computed by from batch IVR. Secondly, we also compare the performance of OIVR and online OLS where OIVR outperforms online OLS in in most cases. In Fig. 3(c) we notice that OIVR with FTRL, ONS, iOGD gives smaller filter error than batch IVR. This is possible since IVR does not explicitly minimize the filter error but instead minimizes the single step prediction error. The consistency result for IVR only holds if the system has truely linear dynamics. However, as our dynamics benchmarks consist of non-linear dynamics, there may exist a linear estimator of the system dynamics that can outperform IVR in terms of minimizing filter error.

College Distance We finally consider an econometrics problem, a traditional application domain for instrumental variable regression, the College Distance vignette. In this experiment, we try to predict future wages given the number of years of education as the explanatory variable (feature) and the distance to the nearest 4 year college as the instrument (Kleiber and Zeileis 2008; Card 1993). The claim is that the distance is correlated with the years of college but is uncorrelated with future wages except through the education level. The goal is to find the meaningful linear coefficient from education level to wages. As such, we do not compare against OLS as it does not try to find a similarly meaningful representation. We see in Table 1 that the online algorithm is able to converge on the solution found in the batch setting.

	IVR	**iOGD**	**ONS**	**OGD**	**FTRL**
Computed A	0.688	0.690	0.689	0.698	0.688

Table 1: Comparison of $\bar{A}$ found using various online no-regret algorithms to the result from batch IVR for the **College Distance** dataset.

Conclusion

We have introduced a new algorithm for Online Instrumental Variable Regression and proved strong theoretical performance bounds with regard to the traditional batch Instrumental Variable Regression setting through a connection to no-regret online algorithms. Through connections between IVR and dynamical system identification, we have introduced a rich new family of online system identification algorithms. Our experimental results show that OIVR algorithms work well in practice on a variety of benchmark datasets.

Acknowledgments

This material is based upon work supported in part by: National Science Foundation Graduate Research Fellowship Grant No. DGE-1252522, DARPA ALIAS contract number HR0011-15-C-0027, and National Science Foundation National Robotics Initiative. The authors also thank Geoff Gordon for valuable and insightful discussions.

References

Abbeel, P., and Ng, A. Y. 2005. Exploration and apprenticeship learning in reinforcement learning. In *ICML*, 1–8. ACM.

Boots, B., and Gordon, G. 2011a. An online spectral learning algorithm for partially observable nonlinear dynamical systems. In *AAAI*.

Boots, B., and Gordon, G. J. 2011b. Predictive state temporal difference learning. In *NIPS*.

Boots, B., and Gordon, G. 2012. Two-manifold problems with applications to nolinear system identification. In *ICML*.

Boots, B.; Siddiqi, S.; and Gordon, G. 2011. Closing the learning planning loop with predictive state representations. *IJRR*.

Bowden, R. J., and Turkington, D. A. 1990. *Instrumental variables*, volume 8. Cambridge University Press.

Cameron, A. C., and Trivedi, P. K. 2005. *Microeconometrics: methods and applications*. Cambridge university press.

Card, D. 1993. Using geographic variation in college proximity to estimate the return to schooling. Technical report, National Bureau of Economic Research.

Cesa-Bianchi, N.; Conconi, A.; and Gentile, C. 2004. On the generalization ability of on-line learning algorithms. *Information Theory, IEEE Transactions on* 50(9):2050–2057.

Greenland, S. 2000. An introduction to instrumental variables for epidemiologists. *International Journal of Epidemiology* 29(4):722–729.

Hardy, G. H. 2000. *Divergent series*. American Mathematical Society.

Hazan, E.; Agarwal, A.; and Kale, S. 2006. Logarithmic regret algorithms for online convex optimization. *Proceedings of the 19th annual conference on Computational Learning Theory (COLT)* 169–192.

Hazan, E., and Kale, S. 2014. Beyond the regret minimization barrier: optimal algorithms for stochastic strongly-convex optimization. *JMLR*.

Hefny, A.; Downey, C.; and J. Gordon, G. 2015. A new view of predictive state methods for dynamical system learning. *arXiv preprint arXiv:1505.05310*.

Kleiber, C., and Zeileis, A. 2008. *Applied econometrics with R*. Springer Science & Business Media.

Kulis, B.; Bartlett, P. L.; Eecs, B.; and Edu, B. 2010. Implicit Online Learning. *Proceedings of the 27th international conference on Machine learning (ICML)* 575–582.

Littlestone, N. 2014. From on-line to batch learning. In *Proceedings of the second annual workshop on Computational learning theory*, 269–284.

Littman, M. L.; Sutton, R. S.; and Singh, S. 2001. Predictive representations of state. In *NIPS*, 1555–1561. MIT Press.

Miguel, E.; Satyanath, S.; and Sergenti, E. 2004. Economic shocks and civil conflict: An instrumental variables approach. *Journal of Political Economy* 112(4):725–753.

Ralaivola, L., and D'Alche-Buc, F. 2004. Dynamical modeling with kernels for nonlinear time series prediction. *NIPS*.

Rao, C. R.; Toutenburg, H.; Shalabh, H. C.; and Schomaker, M. 2008. Linear models and generalizations. *Least Squares and Alternatives (3rd edition) Springer, Berlin Heidelberg New York*.

Research, L. 2015. X-plane. DVD.

Shalev-Shwartz, S. 2011. Online Learning and Online Convex Optimization. *Foundations and Trends in Machine Learning* 4(2):107–194.

Singh, S.; James, M. R.; and Rudary, M. R. 2004. Predictive state representations: A new theory for modeling dynamical systems. In *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence*, UAI '04, 512–519. Arlington, Virginia, United States: AUAI Press.

Söderström, T., and Stoica, P. 2002. Instrumental variable methods for system identification. *Circuits, Systems and Signal Processing* 21(1):1–9.

Wingate, D., and Singh, S. 2006. Kernel predictive linear Gaussian models for nonlinear stochastic dynamical systems. *ICML* 1017–1024.

Zinkevich, M. 2003. Online Convex Programming and Generalized Infinitesimal Gradient Ascent. In *International Conference on Machine Learning (ICML 2003)*, 421–422.

Proofs of Theorems and Lemmas

Proof of Theorem 2

Assume (x_t, y_t, z_t) are i.i.d. and $\mathbf{E}(zz^T)$ is postive definite. Following any online no-regret procedure on the convex L2 losses for M_t, and A_t and computing $\bar{M}_t, \ \bar{A}_t$ as shown in Algorithm 2, we get that:

$$\mathbf{E}\left[\left\|\bar{A}_T M^* z - y\right\|_2^2\right] = \mathbf{E}\left[\left\|A^* M^* z - y\right\|_2^2\right], \ \ T \to \infty \tag{13}$$

$$\bar{A}_T \to A^*, \ \ T \to \infty \tag{14}$$

for the A^*, M^* from Batch IVR *(Alg. 1).*

Proof. Let $\epsilon_t = M^* - \bar{M}_t$. Then,

$$\begin{aligned}\left\|A\bar{M}_t z_t - y_t\right\|_2^2 &= \|A(M^* - \epsilon_t)z_t - y_t\|_2^2 \\ &= \|AM^* z_t - y_t\|_2^2 + \|A\epsilon_t z_t\|_2^2 - 2\left(AM^* z_t - y_t\right)^T \left(A\epsilon_t z_t\right)\end{aligned} \tag{15}$$

Since we run a no-regret online algorithm for A_t on loss function $\|A_t\bar{M}_t z_t - y_t\|_2^2$, we have:

$$\sum_t \|A_t\bar{M}_t z_t - y_t\|_2^2 \leq \text{REGRET}_A + \sum_t \|A^*\bar{M}_t z_t - y_t\|_2^2$$

where $\sum_t$ denotes $\sum_{t=1}^T$.

Applying Eqn. 15 to the right hand side, we get that for any A^* including the minimizer:

$$\sum_t \|A_t\bar{M}_t z_t - y_t\|_2^2 \leq \text{REGRET}_A + \sum_t \|A^* M^* z_t - y_t\|_2^2 + \|A^*\epsilon_t z_t\|_2^2 - 2\left(A^* M^* z_t - y_t\right)^T \left(A^*\epsilon_t z_t\right) \tag{16}$$

Applying Eqn. 15 to the left hand side, we get:

$$\begin{aligned}&\sum_t \|A_t M^* z_t - y_t\|_2^2 + \|A_t\epsilon_t z_t\|_2^2 - 2\left(A_t M^* z_t - y_t\right)^T \left(A_t\epsilon_t z_t\right) \\ &\qquad \leq \text{REGRET}_A + \sum_t \|A^* M^* z_t - y_t\|_2^2 + \|A^*\epsilon_t z_t\|_2^2 - 2\left(A^* M^* z_t - y_t\right)^T \left(A^*\epsilon_t z_t\right)\end{aligned} \tag{17}$$

Rearranging terms,

$$\begin{aligned}\sum_t \|A_t M^* z_t - y_t\|_2^2 &\leq \text{REGRET}_A + \sum_t \|A^* M^* z_t - y_t\|_2^2 + \|A^*\epsilon_t z_t\|_2^2 - 2\left(A^* M^* z_t - y_t\right)^T \left(A^*\epsilon_t z_t\right) \\ &\qquad - \|A_t\epsilon_t z_t\|_2^2 + 2\left(A_t M^* z_t - y_t\right)^T \left(A_t\epsilon_t z_t\right) \qquad (18) \\ &\leq \text{REGRET}_A + \sum_t \|A^* M^* z_t - y_t\|_2^2 + \|A^*\|_F^2 \|\epsilon_t\|_F^2 \|z_t\|_2^2 + \|A_t\|_F^2 \|\epsilon_t\|_F^2 \|z_t\|_2^2 \\ &\qquad + 2\left|\left(A^* M^* z_t - y_t\right)^T \left(A^*\epsilon_t z_t\right)\right| + 2\left|\left(A_t M^* z_t - y_t\right)^T \left(A_t\epsilon_t z_t\right)\right| \qquad (19)\end{aligned}$$

The Cauchy-Swartz and the triangle inequality gives us that for any A:

$$\begin{aligned}\left|\left(AM^* z_t - y_t\right)^T \left(A\epsilon_t z_t\right)\right| &\leq \|AM^* z_t - y_t\|_2 \|A\epsilon_t z_t\|_2 \\ &\leq \left(\|AM^* z_t\|_2 + \|y_t\|_2\right) \|A\epsilon_t z_t\|_2 \\ &\leq \left(\|A\|_F \|M^*\|_F \|z_t\|_2 + \|y_t\|_2\right) \|A\|_F \|\epsilon_t\|_F \|z_t\|_2 \\ &\leq \left(\|A\|_F^2 \|z_t\|_2^2 \|M^*\|_F + \|A\|_F \|z_t\|_2 \|y_t\|_2\right) \|\epsilon_t\|_F\end{aligned} \tag{20}$$

Assuming that $\|z_t\|_2$, $\|y_t\|_2$, $\|M^*\|_F$, $\|A_t\|_F$, $\|A^*\|_F$ are each always upper bounded by some positive constant, define positive constants C_1 and C_2 such that:

$$C_1 \geq \|z_t\|_2^2 \left(\|A_t\|_F^2 + \|A^*\|_F^2\right) \tag{21}$$

$$C_2 \geq 2\left(\|A_t\|_F^2 \|z_t\|_2^2 \|M^*\|_F + \|A_t\|_F \|z_t\|_2 \|y_t\|_2 + \|A^*\|_F^2 \|z_t\|_2^2 \|M^*\|_F + \|A^*\|_F \|z_t\|_2 \|y_t\|_2\right) \tag{22}$$

Utilizing Eqn. 20 with Eqn. 19:

$$\sum_t \|A_t M^* z_t - y_t\|_2^2 \leq \textsc{Regret}_A + \sum_t \|A^* M^* z_t - y_t\|_2^2 + C_1 \|\epsilon_t\|_F^2 + C_2 \|\epsilon_t\|_F \tag{23}$$

Let $\mathbf{E}$ denote the expectation with regards to the whole sequence of data. Assuming that z_t, x_t, and y_t are i.i.d. we get:

$$\mathbf{E}\left[\sum_t \|A_t M^* z - y\|_2^2\right] \leq \mathbf{E}\left[\textsc{Regret}_A\right] + \mathbf{E}\left[\sum_t \|A^* M^* z - y\|_2^2\right] + \mathbf{E}\left[\sum_t \Big(C_1 \|\epsilon_t\|_F^2 + C_2 \|\epsilon_t\|_F\Big)\right] \tag{24}$$

Now let us consider how $\|\epsilon_t\|_F$ can be upper bounded. For ϵ_t, since we run a no-regret online algorithm on loss $\|M_t z_t - x_t\|_2^2$ assuming z_t, x_t, and y_t are i.i.d and that $\mathbf{E}[zz^T]$ is positive definite, we use Lemma 1 to get:

$$\mathbf{E}\|\epsilon_t\|_F^2 = \mathbf{E}\left\|\bar{M}_t - M^*\right\|_F^2 \leq \frac{1}{t} r_M(t) \to 0, \text{ as } t \to \infty \tag{25}$$

$$\Rightarrow \mathbf{E}\|\epsilon_t\|_F = \sqrt{(\mathbf{E}\|\epsilon_t\|_F)^2} \leq \sqrt{\mathbf{E}\|\epsilon_t\|_F^2} \leq \sqrt{\frac{1}{t} r_M(t)} \to 0, \text{ as } t \to \infty \tag{26}$$

We get the inequality in Eqn. 26 since $\mathbf{Var}(\epsilon_t) = \mathbf{E}\left[\epsilon_t^2\right] - \mathbf{E}\left[\epsilon_t\right]^2 \geq 0$. Dividing by T in Eqn. 24, we get:

$$\frac{1}{T}\mathbf{E}\left[\sum_t \|A_t M^* z - y\|_2^2\right] \leq \mathbf{E}\left[\frac{\textsc{Regret}_A}{T}\right] + \mathbf{E}\left[\frac{1}{T}\sum_t \|A^* M^* z - y\|_2^2\right] + \mathbf{E}\left[\frac{1}{T}\sum_t \left(C_1 \frac{r_M(t)}{t} + C_2 \sqrt{\frac{r_M(t)}{t}}\right)\right]$$

Note that for no-regret online algorithms, $\lim_{t\to\infty} r_M(t)/t = 0$ since $r_M(t) \in o(T)$. Then because $\mathbf{E}\left[\frac{1}{T}\sum_t \|A^* M^* z - y\|_2^2\right] = \mathbf{E}\left[\|A^* M^* z - y\|_2^2\right]$, applying Cesaro Mean (Hardy 2000) and taking the limit of T, we get:

$$\frac{1}{T}\mathbf{E}\left[\sum_t \|A_t M^* z - y\|_2^2\right] \leq \mathbf{E}\left[\|A^* M^* z - y\|_2^2\right], \; T \to \infty \tag{27}$$

Thus, we have shown the algorithm is no-regret. Let $\bar{A}_T = \frac{1}{T}\sum_t A_t$. Using Jensen's inequality, we get:

$$\mathbf{E}\left[\left\|\bar{A}_T M^* z - y\right\|_2^2\right] \leq \frac{1}{T}\mathbf{E}\left[\sum_t \|A_t M^* z - y\|_2^2\right] \leq \mathbf{E}\left[\|A^* M^* z - y\|_2^2\right], \; T \to \infty \tag{28}$$

Since the above is valid for any A^*, let $A^* = \arg\min_A \mathbf{E}\left[\|A M^* z - y\|_2^2\right]$. Then by construction,

$$\mathbf{E}\left[\left\|\bar{A}_T M^* z - y\right\|_2^2\right] \geq \mathbf{E}\left[\|A^* M^* z - y\|_2^2\right], \; T \to \infty \tag{29}$$

Therefore, by Eqn. 28 and Eqn. 29, we have that:

$$\mathbf{E}\left[\left\|\bar{A}_T M^* z - y\right\|_2^2\right] \to \mathbf{E}\left[\|A^* M^* z - y\|_2^2\right], \; T \to \infty \tag{30}$$

With $\mathbf{E}\left[zz^T\right] \succ 0$ resulting in a strongly convex objective, we get a unique minimizer for the objective. Therefore,

$$\bar{A}_T \to A^*, \; T \to \infty \tag{31}$$

Hence, we prove the theorem.

□

Proof of Lemma 1

Proof. Since we use no-regret algorithms on losses $\{\ell_t\}$, for any β^*, we have:

$$\sum_t \ell_t(\beta_t) - \sum_t \ell_t(\beta^*) \leq r(T) \in o(T), \tag{32}$$

where $\sum_t$ denotes $\sum_{t=1}^T$ and $r(T)$ is the regret rate of the online algorithm (upper bound of the regret), which is sublinear for all no-regret online algorithms. Taking the expectation of l_t on both sides of the equation (let us assume $\ell = \mathbf{E}_{\ell_t}(\ell_t), \forall t$), we have:

$$\mathbf{E}\left[\sum_t \ell(\beta_t) - \sum_t \ell(\beta^*)\right] \in o(T). \tag{33}$$

The expectation taken here is of β_t with respect to the stochastic losses so far. β_t (the parameter being optimized) is a random variable that depends on only the previous $t-1$ loss functions. Thus, the loss ℓ_t at step t is independent of β_t. The expectation over the losses becomes $\mathbf{E}_{\ell_1,\ldots\ell_t}[\ell_t(\beta_t)] = \mathbf{E}_{\ell_1,\ldots\ell_{t-1}}[\ell(\beta_t)]$, where the expectation of ℓ_t is ℓ (i.e. the loss is stochastic due to drawing the next set of data points from an i.i.d. distribution).

Since ℓ is convex with respect to β, from Jensen's Inequality, we have:

$$\ell(\frac{1}{T}\sum_t \beta_t) \le \frac{1}{T}\sum_t \ell(\beta_t). \tag{34}$$

Combining Eqn. 34 and Eqn. 33, we have:

$$\mathbf{E}\left[\ell(\bar{\beta}) - \ell(\beta^*)\right] \le \mathbf{E}\left[\frac{1}{T}\sum_t \ell(\beta_t) - \frac{1}{T}\sum_t \ell(\beta^*)\right] \le \mathbf{E}\left[\frac{1}{T} r(T)\right] \to 0, \text{ as } T \to \infty. \tag{35}$$

When ℓ is a α strongly-convex loss function with respect to β under norm $\|\cdot\|$ and $\beta^* = \arg\min_\beta \ell(\beta)$, we have that the convergence in the objective gives us convergence in the parameter:

$$\frac{\alpha}{2}\left\|\bar{\beta} - \beta^*\right\| \le \ell(\bar{\beta}) - \ell(\beta)^* \le \frac{1}{T} r(T). \tag{36}$$

Putting the expectation back and taking T to infinity, we have:

$$\mathbf{E}\left[\left\|\bar{\beta} - \beta^*\right\|\right] \le \frac{2}{\alpha T} r(T) \to 0, \text{ as } T \to \infty. \tag{37}$$

□