

A Review of Statistical Guarantees for the EM Algorithm with Simulation Examples

Zhi Zhang*

1 Introduction

It is very challenging to compute the maximum likelihood estimate (MLE) in most of the incomplete data problems setting. The expectation-maximization (EM) algorithm is a very useful tool to solve such problems. But there is a gap between the practical use of EM and its theoretical properties. In most models, the EM algorithm is only guaranteed to return a local optimum of the sample likelihood function without good statistical properties of MLE. The main contribution of the paper[1] is to provide a guarantee of convergence to a local optimum that matches the performance of MLE under specific conditions. The authors achieve that by two steps: first derive the statistical guarantees of the EM and first-order EM algorithms at the population level, then extend the result to the case of finite samples. In addition, the authors present three canonical problems which satisfy the conditions of the theorems and provide a tight characterizations of the region of convergence for them: symmetric mixture of two Gaussians, symmetric mixture of two regressions and linear regression with missing completely at random.

To be specific, the authors make advances over the classical results in the following three specific directions:

1. By imposing conditions of gradient smoothness, strong concavity and smoothness on the auxiliary Q-function underlying the EM algorithm, they give a quantitative characterization of the region of convergence to the population global optimum. The existing research can only guarantee convergence to a fixed point when initialized around its some neighborhood but can't quantify the size of the neighborhood.
2. The authors are the first one who analyze the EM algorithm on the population level and then quantify the minimal requirement of sample size to avoid spurious fixed points far away from the population MLE. The classical results on the EM algorithm are all based on the finite sample size and can not make sure that any fixed points of the sample likelihood are close to the population MLE.
3. A precise characterization of initialization is presented in the paper. A two-stage estimator, which involves the initialization of the moments method and the refinement of the EM algorithm, performs well empirically. The authors' theoretical results help explain this behavior and furthermore provide guideline on the refinement stage.

Apart from the contributions mentioned above, the authors' effort in the treatment of three examples is also remarkable. They show that their conditions hold in a large region around the

*Department of Statistics, UC Davis, wwzzhang@ucdavis.edu. This is the project report for STA243 Computational Statistics

MLE, and that the size of this region is determined by interpretable problem-dependent quantities. Extensive simulations are carried out to confirm the theorems.

1.1 Related work

The EM algorithm has a long history (e.g., [2], [3], [5], [6], [11], [12], [13]) and its modern general form was introduced by Dempster, Laird and Rubin [4], who also established its well-known monotonicity properties. Wu [14] provided guarantee for the EM algorithm to converge to the unique global optimum when the likelihood is unimodal and certain regularity conditions hold. When the likelihood function is multi-modal, which is more common in reality, existing works can only guarantee convergence to some local optimum of the likelihood at an asymptotically geometric rate (see, e.g., [7], [8], [9], [10]). This type guarantee is not enough to promise convergence to a “good” local optimum. The local optimum can be far away from any global optimum of the likelihood. The paper closes the gap by guaranteeing geometric convergence to a “good” EM fixed point.

2 Methodology

The main methodological contributions of the paper is the construction of the theorems on the population EM and first-order EM algorithms, which are then used on proof the convergence theorems of their finite-sample counterparts. This innovation is very intuitive. In order to show that the algorithm can converge to a global optimal point, we must know the existence of the optimal point and have an insight on the perfect situation, e.g., when we have infinite samples. This intuition makes it natural to introduce the population EM and first-order EM algorithms. If the population EM and first-order EM algorithms perform well, people will want to know the performance of their finite-sample counterparts, which is of more interest in reality. Another intuition is to bound the difference between the sample and population function or gradient, which is called an empirical process. When the empirical process can be bounded well, which will be achieved by sufficiently large sample size, it can be expected that the behavior of the finite-sample EM and first-order EM algorithm will be very similar to their population counterparts. By this way, the authors arrived their ultimate interest by building theorems on the finite-sample EM and first-order EM algorithm.

To be more specific, at first, the author gives conditions under which the population algorithms are contractive to the global optimal point, e.g., the true parameter θ^* , when initialized in a ball around the θ^* . These conditions allow us to establish the region of attraction of θ^* .

For the first-order EM algorithms, they exploit the oracle auxiliary function $q(\theta)$, whose convergence at a geometric rate to the global optimum is guaranteed under some standard regularity conditions by classical theory on gradient methods. Then they impose the condition of gradient smoothness in the neighborhood of θ^* to make sure that the the gradient of the auxiliary function $Q(\theta|\theta)$ with respect to its first argument is close enough to $\nabla q(\theta)$. So the convergence result on $q(\theta)$ can be extended to $Q(\theta|\theta)$. Furthermore, provided that the sample size is large enough, the sample gradient $\nabla Q_n(\theta|\theta)$ will be within a small area of $\nabla Q(\theta|\theta)$ with high probability. This fact ensures the convergence properties of the empirical EM algorithm.

For the EM algorithm, the authors first define the operator $M(\theta) = \arg \max_{\theta' \in \Omega} Q(\theta|\theta')$. Then, conditioned on the first-order stability (FOS) of $Q(\cdot|\theta)$, the KKT conditions, and a similar globally strong concavity of the function $q(\theta)$, the population EM operator M is contractive over a Euclidean ball of θ^* . Then by controlling the quantity of the empirical process $M(\theta) - M_n(\theta)$ in the neighborhood of θ^* with high probability, which can be achieved by sufficiently large sample size, the authors prove the convergence of the EM iterates to the global optimal point, e.g., the true parameters.

The authors make a remark that the first order EM algorithm is identical to the gradient ascent on the marginal log-likelihood function. The intuition behind the analysis is the exploiting of the Q-function, which makes it possible to prove guarantees for a specific class of models whose log likelihood is not concave but the Q-function is concave and satisfies some conditions.

What is of interest is the convergence property of sample-based EM algorithm. However, it is very difficult to link the estimate $\hat{\theta}$ returned by the algorithm to θ^* directly. The authors' solution is very intuitive and clear - step by step, which goes from $q(\theta)$ to $Q(\theta|\theta)$, finally to $Q_n(\theta|\theta)$. The proof of these theorems is not difficult, while the proof of the corollaries involve several techniques such as symmetrization, contraction and concentration inequalities, which are common in the field of high dimensional statistics. This is because when apply the theorems to the canonical problems, e.g., to get the corollaries, a large part of the technical effort is devoted to establishing bounds on the empirical process $\{\nabla Q(\theta|\theta) - \nabla Q_n(\theta|\theta), \theta \in \mathbb{B}_2(r; \theta^*)\}$.

Other methodological progresses are attributed to the corollaries of the three canonical problems. For instance, for the Gaussian mixture models, the authors use a simple setting with two components, balanced weights and isotropic covariances. Then the goal is to estimate the unknown mean vector θ^* . The difficulty of estimating such a mixture model can be characterized by the signal-to-noise ratio $\frac{\|\theta^*\|_2}{\sigma}$ (SNR). Finally, the corollary of population result for the first-order EM algorithm for Gaussian mixtures is built after verifying that the Gaussian mixture model satisfies the gradient smoothness, λ -strong concavity and μ -smoothness. The condition on SNR is intuitive and has empirical support.

3 Theoretical Results

Let Y and Z be random variables in the sample spaces \mathcal{Y} and \mathcal{Z} with joint density function f_{θ^*} that belongs to some parameterized family $\{f_\theta | \theta \in \Omega\}$, where Ω is some nonempty convex set of parameters. Here Y is the observed component and Z is the latent structure in the data. For each $\theta \in \Omega$, denote $k_\theta(z|y)$ to be the conditional density of z given y . Suppose that we have n i.i.d. observations $\{y_i\}_{i=1}^n$, the standard EM algorithm maximizes the log-likelihood function $\ell_n(\theta) = \frac{1}{n} \sum_{i=1}^n \log [\int_{\mathcal{Z}} f_\theta(y_i, z) dz]$ by maximizing the finite-sample Q function $Q_n(\theta|\theta')$ at $\theta' = \theta^t$.

$$\theta^{t+1} = \arg \max_{\theta \in \Omega} Q_n(\theta|\theta^t) \quad \text{where} \quad Q_n(\theta|\theta') = \frac{1}{n} \sum_{i=1}^n \left(\int_{\mathcal{Z}} k_{\theta'}(z|y_i) \log f_\theta(y_i, z) dz \right) \quad (1)$$

Definition 3.1. Denote θ^* to be the true parameter and g_{θ^*} to be the marginal density of the observed samples. Then the population counterpart of EM algorithm maximizes the log-likelihood function

$$\ell(\theta) = \int_{\mathcal{Y}} \log \left[\int_{\mathcal{Z}} f_\theta(y, z) dz \right] g_{\theta^*}(y) dy \quad (2)$$

by maximizing the population Q-function given below in each iterations at $\theta' = \theta^t$

$$Q(\theta|\theta') = \int_{\mathcal{Y}} \left(\int_{\mathcal{Z}} k_{\theta'}(z|y) \log f_\theta(y, z) dz \right) g_{\theta^*}(y) dy \quad (3)$$

The main theoretical results of the paper can be viewed as three parts: the convergence theorems of the standard EM and the first order EM algorithm (Theorem 1 and 4 in the paper), their counterparts of the sample-based algorithms (Theorem 2 and 5(a)) and the extension of sample-splitting algorithms (Theorem 3 and 5(b)), and the corollaries which apply these theorems to the specific three canonical problems of interest. In particular, the update rule for three versions for first order EM algorithm is given by Table 1 and there are mainly three common conditions that need to be guaranteed in these theorems.

Algorithm	Update Rules
Population-level first-order EM algorithm	$\theta^{t+1} = \theta^t + \alpha \nabla Q(\theta \theta^t) \Big _{\theta=\theta^t}$
Sample-based first-order EM algorithm	$\theta^{t+1} = \theta^t + \alpha \nabla Q_n(\theta \theta^t) \Big _{\theta=\theta^t}$
Sample-splitting first order EM algorithm	$\theta^{t+1} = \theta^t + \alpha \nabla Q_{\lfloor n/T \rfloor}(\theta \theta^t) \Big _{\theta=\theta^t}$

Table 1: The update rule for three first-order EM algorithm. The gradient $\nabla Q_n(\theta|\theta^t)$ is taken with respect to the first argument of the Q -function and $\nabla Q_{\lfloor n/T \rfloor}$ denotes the finite sample Q -function computed using a fresh subset of $\lfloor n/T \rfloor$ samples at each iteration.

Condition 3.1 (Gradient smoothness). For an appropriately small parameter $\gamma \geq 0$, we have that

$$\|\nabla q(\theta) - \nabla Q(\theta|\theta)\|_2 \leq \gamma \|\theta - \theta^*\|_2 \quad \text{for all } \theta \in \mathbb{B}_2(r; \theta^*) \quad (4)$$

Condition 3.2 (λ -strong concavity). There is some $\lambda > 0$ such that for all pairs $\theta_1, \theta_2 \in \mathbb{B}_2(r; \theta^*)$

$$q(\theta_1) - q(\theta_2) - \langle \nabla q(\theta_2), \theta_1 - \theta_2 \rangle \leq -\frac{\lambda}{2} \|\theta_1 - \theta_2\|_2^2 \quad (5)$$

When we require this condition to hold for all pairs $\theta_1, \theta_2 \in \Omega$ we refer to this as global λ -strong concavity.

Condition 3.3 (μ -smoothness). There is some $\mu > 0$ such that for all $\theta_1, \theta_2 \in \mathbb{B}_2(r; \theta^*)$

$$q(\theta_1) - q(\theta_2) - \langle \nabla q(\theta_2), \theta_1 - \theta_2 \rangle \geq -\frac{\mu}{2} \|\theta_1 - \theta_2\|_2^2 \quad (6)$$

- For theorem 1-3, there are some common parameters. They are **triplet** (γ, λ, μ) such that $0 \leq \gamma < \lambda \leq \mu$, **radius** $r > 0$, **step size** $\alpha = \frac{2}{\mu+\lambda}$ and **initialization** $\theta^0 \in \mathbb{B}_2(r; \theta^*)$.
- Theorem 1 is related to the analysis on **population level for first order EM algorithm**. This theorem guarantees a **geometric rate of convergence** toward θ^* for population level EM algorithm.
- Theorem 2 is related to the convergence rate for **sample-based first-order EM algorithm**. The extra parameter $\varepsilon_Q^{\text{unif}}(n, \delta)$ is the smallest scalar such that $\mathbb{P}\left(\sup \mathcal{E} \leq \varepsilon_Q^{\text{unif}}(n, \delta)\right) \geq 1 - \delta$ where the empirical process is defined as $\mathcal{E} = \{\nabla Q(\theta|\theta) - \nabla Q_n(\theta|\theta), \theta \in \mathbb{B}_2(r; \theta^*)\}$.
- Theorem 3 is related to the convergence rate for **mini-batch sample-based first-order EM algorithm**. The full data is divided into T subsets of size $\lfloor n/T \rfloor$. The extra parameter $\varepsilon_Q(n, \delta)$ be the smallest scalar such that for any fixed $\theta \in \mathbb{B}_2(r; \theta^*)$,

$$\mathbb{P}\left[\left\|\nabla Q_n(\theta|\theta^t)\Big|_{\theta=\theta^t} - \nabla Q(\theta|\theta^t)\Big|_{\theta=\theta^t}\right\|_2 < \varepsilon_Q(n, \delta)\right] \geq 1 - \delta$$

The theoretical results of theorem 1-3 are summarized in Table 2. The **more detailed proof** than the paper of Theorem 1 is provided in Appendix A.1 and the proof of theorem 2 and theorem 3 are provided in Appendix A.2. Among these three theorems, conditions 3.1 through 3.3 hold over the ball $\mathbb{B}_2(r; \theta^*)$ for a reasonably large choice of r . In practice, there are some popular examples that satisfies the conditions in Theorem 1, Theorem 2 and Theorem 3. The author applies these theorems on three classes of statistical models for which the EM algorithm is frequently applied, namely, Gaussian mixture models, mixtures of regressions and regression with missing covariates. The details for theoretical illustration are provided in Appendix B. For this three specific models, the authors check the conditions we discuss above and the applies three theorems

Algorithm	Extra parameter	Convergence rate
Population	No extra parameter	$\ \theta^t - \theta^*\ _2 \leq \left(1 - \frac{2\lambda-2\gamma}{\mu+\lambda}\right)^t \ \theta^0 - \theta^*\ _2$
Sample-based	$\varepsilon_Q^{\text{unif}}(n, \delta) \leq (\lambda - \gamma)r$	$\ \theta^t - \theta^*\ _2 \leq \left(1 - \frac{2\lambda-2\gamma}{\mu+\lambda}\right)^t \ \theta^0 - \theta^*\ _2 + \frac{\varepsilon_Q^{\text{unif}}(n, \delta)}{\lambda - \gamma}$
Sample-splitting	$\varepsilon_Q\left(\frac{n}{T}, \frac{\delta}{T}\right) \leq (\lambda - \gamma)r$	$\ \theta^t - \theta^*\ _2 \leq \left(1 - \frac{2\lambda-2\gamma}{\mu+\lambda}\right)^t \ \theta^0 - \theta^*\ _2 + \frac{\varepsilon_Q(n/T, \delta/T)}{\lambda - \gamma}$

Table 2: Summary Table for population and sample-based EM algorithm. For a specific tolerance parameter $\delta > 0$, the conditions for extra parameter in sample-based and sample splitting holds with probability at least $1 - \delta$.

to obtain the corresponding convergence results. In these examples, the strong concavity and smoothness conditions are straightforward, whereas establishing gradient smoothness (Condition 3.1) is more challenging. Establishing that the gradient condition holds over (nearly) optimally-sized regions involves carefully leveraging properties of the generative model as well as smoothness properties of the log-likelihood function.

Besides, the author also provides with some extensions for first order EM algorithm both on population level and sample based methods. For population level, the authors establish the convergence rate for EM algorithm. For sample-based method, the authors provide some general guarantees for sample-based EM algorithm. These theorems are provided in Appendix A.3 and are quite similar to the main results above (Theorem 1-3).

4 Experimental Details

4.1 Setup

We set up the simulation experiments according to the three problems introduced in the paper (Mixture of Gaussians, Mixture of Regressions, and Missing Data Regression). We implemented the EM updates and first-order EM updates for each of problems, in a total, there are 3×2 (problem, algorithm) combinations.

For each problem-algorithm combination, we examined the logarithm statistical error $\|\theta^t - \theta^*\|_2$ and optimization error $\|\theta^t - \hat{\theta}\|_2$ over iterations. We replicated with 10 times, and fixed the dimension $d = 10$, sample size $n = 1000$, sigma $\sigma = 1$, and signal-to-noise ratio $\frac{\|\theta^*\|_2}{\sigma} = 2$. In addition, while each replicate we randomly did the initialization, but we also made sure that the θ^0 was in the range of radius of convergence stated by the theorems. For missing data problem, we let $p = 0.2$ for every sample's j th dimension to be randomly missing. The code was attached in the Appendices.

4.2 Results

Fig 1 (a,c,e) shows the EM updates for three problems, (b, d, f) shows First-Order EM updates. We observe that the statistical error (red curve represented) decreases geometrically, and then level off at a plateau, while the optimization error (blue curve) keeps decreasing geometrically. The results also show that the first-order EM takes more iteration to converge. Besides, we notice that the Missing data regression has more variance compared to other two problems, this might be due to randomness introduced during the data creation makes the problem naturally harder. This randomness might also cause the statistical error fluctuation as decreasing, then increasing and eventually plateau in the first-order EM. Figure 1 (g) confirms that the convergence rate increases as the increasing of the signal-to-noise ratio.

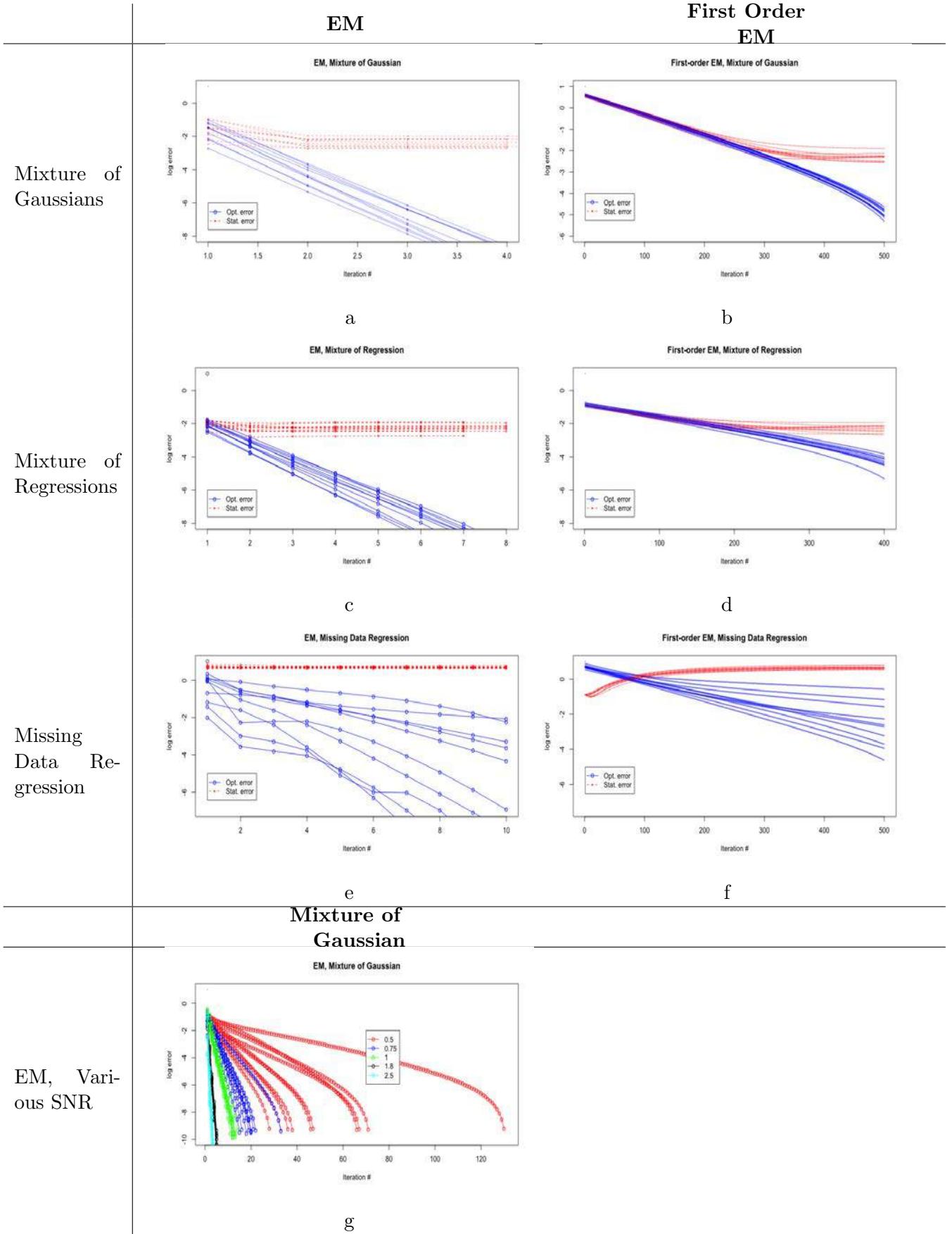


Figure 1: The full sets of replication results for figure 5,6,7,8 in the Balakrishnan's paper

References

- [1] Sivaraman Balakrishnan, Martin J. Wainwright, and Bin Yu. Statistical guarantees for the em algorithm: From population to sample-based analysis. *Ann. Statist.*, 45(1):77–120, 02 2017.
- [2] Leonard E Baum, Ted Petrie, George Soules, and Norman Weiss. A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains. *The annals of mathematical statistics*, 41(1):164–171, 1970.
- [3] Evelyn ML Beale and Roderick JA Little. Missing values in multivariate analysis. *Journal of the Royal Statistical Society: Series B (Methodological)*, 37(1):129–145, 1975.
- [4] Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22, 1977.
- [5] Herman O Hartley. Maximum likelihood estimation from incomplete data. *Biometrics*, 14(2):174–194, 1958.
- [6] Michael Healy and Michael Westmacott. Missing values in experiments analysed on automatic computers. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 5(3):203–206, 1956.
- [7] Alfred O Hero and Jeffrey A Fessler. Convergence in norm for alternating expectation-maximization (em) type algorithms. *Statistica Sinica*, pages 41–54, 1995.
- [8] Geoffrey J McLachlan and Thriyambakam Krishnan. *The EM algorithm and extensions*, volume 382. John Wiley & Sons, 2007.
- [9] Xiao-Li Meng et al. On the rate of convergence of the ecm algorithm. *The Annals of Statistics*, 22(1):326–339, 1994.
- [10] Xiao-Li Meng and Donald B Rubin. On the global and componentwise rates of convergence of the em algorithm. *Linear Algebra and its Applications*, 199:413–425, 1994.
- [11] T Orchard and MA Woodbury. A missing information principle: theory and applications proceedings of the 6th berkeley symposium on mathematical statistics vol, 1972.
- [12] Donald B Rubin. Characterizing the estimation of parameters in incomplete-data problems. *Journal of the American Statistical Association*, 69(346):467–474, 1974.
- [13] Rolf Sundberg. Maximum likelihood theory for incomplete data from an exponential family. *Scandinavian Journal of Statistics*, pages 49–58, 1974.
- [14] CF Jeff Wu. On the convergence properties of the em algorithm. *The Annals of statistics*, pages 95–103, 1983.

A Additional Proof Details

A.1 Additional Proof Details for Population-level Analysis

Lemma A.1. For a function q with the λ -strong concavity and μ -smoothness properties (Conditions 3.2 and 3.3), the oracle iterates

$$\tilde{\theta}^{t+1} = \tilde{\theta}^t + \alpha \nabla q(\tilde{\theta}^t) \quad \text{for } t = 0, 1, 2, \dots$$

with stepsize $\alpha = \frac{2}{\mu+\lambda}$ are linearly convergent:

$$\|\theta^t + \alpha \nabla q(\theta)|_{\theta=\theta^t} - \theta^*\|_2 \leq \left(\frac{\mu - \lambda}{\mu + \lambda} \right) \|\theta^t - \theta^*\|_2 \quad (7)$$

Proof. Note that q is λ -strong concavity and μ -smooth, we see that $f = -q$ is λ -strong convex and μ -smooth. Therefore, we define

$$h(\theta) = f(\theta) - \frac{\lambda}{2} \|\theta\|_2^2$$

and we have

$$\begin{aligned} 0 &\leq (\nabla h(\theta_1) - \nabla h(\theta_2))^\top (\theta_1 - \theta_2) \\ &= (\nabla f(\theta_1) - \nabla f(\theta_2))^\top (\theta_1 - \theta_2) - \lambda \|\theta_1 - \theta_2\|_2^2 \\ &\leq (\mu - \lambda) \|\theta_1 - \theta_2\|_2^2 \end{aligned}$$

Therefore, co-coercivity of $\nabla h(\theta)$ can be written as

$$(\nabla f(\theta_1) - \nabla f(\theta_2))^\top (\theta_1 - \theta_2) \geq \frac{\lambda\mu}{\lambda + \mu} \|\theta_1 - \theta_2\|_2^2 + \frac{1}{\lambda + \mu} \|\nabla f(\theta_1) - \nabla f(\theta_2)\|_2^2$$

or equivalently

$$(\nabla q(\theta_1) - \nabla q(\theta_2))^\top (\theta_1 - \theta_2) \leq -\frac{\lambda\mu}{\lambda + \mu} \|\theta_1 - \theta_2\|_2^2 - \frac{1}{\lambda + \mu} \|\nabla q(\theta_1) - \nabla q(\theta_2)\|_2^2$$

Now, following analysis indicates the desired results

$$\begin{aligned} \|\theta^t + \alpha \nabla q(\theta^t) - \theta^*\|_2^2 &= \|\theta^t - \theta^*\|_2^2 + 2\alpha \nabla q(\theta^t)^\top (\theta^t - \theta^*) + \alpha^2 \|\nabla q(\theta^t)\|_2^2 \\ &= \|\theta^t - \theta^*\|_2^2 + 2\alpha (\nabla q(\theta^t) - \nabla q(\theta^*))^\top (\theta^t - \theta^*) + \alpha^2 \|\nabla q(\theta^t) - \nabla q(\theta^*)\|_2^2 \\ &\leq \|\theta^t - \theta^*\|_2^2 - 2\alpha \left(\frac{\lambda\mu}{\lambda + \mu} \|\theta^t - \theta^*\|_2^2 + \frac{1}{\lambda + \mu} \|\nabla q(\theta^t) - \nabla q(\theta^*)\|_2^2 \right) \\ &\quad + \alpha^2 \|\nabla q(\theta^t) - \nabla q(\theta^*)\|_2^2 \\ &= \left(1 - 2\alpha \frac{\lambda\mu}{\lambda + \mu} \right) \|\theta^t - \theta^*\|_2^2 + \left(\alpha^2 - \frac{2\alpha}{\lambda + \mu} \right) \|\nabla q(\theta^t) - \nabla q(\theta^*)\|_2^2 \\ &= \left(1 - \frac{4\lambda\mu}{(\lambda + \mu)^2} \right) \|\theta^t - \theta^*\|_2^2 = \left(\frac{\mu - \lambda}{\mu + \lambda} \right)^2 \|\theta^t - \theta^*\|_2^2 \end{aligned}$$

□

Proof of Theorem 1. By definition of the first-order EM update

$$\theta^{t+1} = \theta^t + \alpha \nabla Q(\theta|\theta^t) \Big|_{\theta=\theta^t}, \quad \text{for } t = 0, 1, 2, \dots$$

we have

$$\begin{aligned}
\|\theta^t + \alpha \nabla Q(\theta|\theta^t)|_{\theta=\theta^t} - \theta^*\|_2 &= \|\theta^t + \alpha \nabla q(\theta)|_{\theta=\theta^t} - \alpha \nabla q(\theta)|_{\theta=\theta^t} + \alpha \nabla Q(\theta|\theta^t)|_{\theta=\theta^t} - \theta^*\|_2 \\
&\leq \|\theta^t + \alpha \nabla q(\theta)|_{\theta=\theta^t} - \theta^*\|_2 + \alpha \|\nabla q(\theta)|_{\theta=\theta^t} - \nabla Q(\theta|\theta^t)|_{\theta=\theta^t}\|_2 \\
&\leq \left(\frac{\mu - \lambda}{\mu + \lambda} \right) \|\theta^t - \theta^*\|_2 + \alpha \gamma \|\theta^t - \theta^*\|_2
\end{aligned}$$

The first inequality holds by triangle inequality and the second inequality holds by Lemma A.1 and Condition 3.1. Substituting $\alpha = \frac{2}{\mu + \lambda}$ and performing some algebra yields the claim.

$$\begin{aligned}
\|\theta^t + \alpha \nabla Q(\theta|\theta^t)|_{\theta=\theta^t} - \theta^*\|_2 &\leq \left(\frac{\mu - \lambda}{\mu + \lambda} \right) \|\theta^t - \theta^*\|_2 + \alpha \gamma \|\theta^t - \theta^*\|_2 \\
&= \left(\frac{\mu - \lambda}{\mu + \lambda} \right) \|\theta^t - \theta^*\|_2 + \frac{2\gamma}{\lambda + \mu} \|\theta^t - \theta^*\|_2 \\
&= \frac{2\gamma + \mu - \lambda}{\lambda + \mu} \|\theta^t - \theta^*\|_2 = \left(1 - \frac{2\lambda - 2\gamma}{\mu + \lambda} \right) \|\theta^t - \theta^*\|_2 \\
&\leq \left(1 - \frac{2\lambda - 2\gamma}{\mu + \lambda} \right)^t \|\theta^0 - \theta^*\|_2
\end{aligned}$$

□

A.2 Additional Proof Details for Sample-based Analysis

Proof of Theorem 2. With probability at least $1 - \delta$ we have that for any $\theta^s \in \mathbb{B}_2(r; \theta^*)$

$$\|\nabla Q_n(\theta|\theta^s)|_{\theta=\theta^s} - \nabla Q(\theta|\theta^s)|_{\theta=\theta^s}\|_2 \leq \varepsilon_Q^{\text{unif}}(n, \delta) \quad (8)$$

We perform the remainder of our analysis under this event. Defining $\kappa = \left(1 - \frac{2\lambda - 2\gamma}{\lambda + \mu}\right)$, we make following claim

Claim A.1. For each iteration $s \in \{0, 1, 2, \dots\}$,

$$\|\theta^{s+1} - \theta^*\|_2 \leq \kappa \|\theta^s - \theta^*\|_2 + \alpha \varepsilon_Q^{\text{unif}}(n, \delta) \quad (9)$$

We prove this claim by induction on the iteration number. The base case is $s = 0$. In this case, we have

$$\begin{aligned}
\|\theta^1 - \theta^*\|_2 &= \|\theta^0 + \alpha \nabla Q_n(\theta|\theta^0)|_{\theta=\theta^0} - \theta^*\|_2 \\
&\leq \|\theta^0 + \alpha \nabla Q(\theta|\theta^0)|_{\theta=\theta^0} - \theta^*\|_2 + \alpha \|\nabla Q(\theta|\theta^0)|_{\theta=\theta^0} - \nabla Q_n(\theta|\theta^0)|_{\theta=\theta^0}\|_2 \\
&\leq \kappa \|\theta^0 - \theta^*\|_2 + \alpha \varepsilon_Q^{\text{unif}}(n, \delta)
\end{aligned}$$

The first inequality holds by triangle inequality and the second inequality holds by Theorem 1. By condition $\varepsilon_Q^{\text{unif}}(n, \delta) \leq (\lambda - \gamma)r$, we have

$$\begin{aligned}
\|\theta^1 - \theta^*\|_2 &\leq \kappa \|\theta^0 - \theta^*\|_2 + \alpha \varepsilon_Q^{\text{unif}}(n, \delta) \\
&\leq \left(1 - \frac{2\lambda - 2\gamma}{\lambda + \mu}\right) r + \frac{2(\lambda - \gamma)r}{\lambda + \mu} = r
\end{aligned}$$

This implies $\theta^1 \in \mathbb{B}_2(r; \theta^*)$. In the induction from $s \mapsto s + 1$, suppose that $\|\theta^s - \theta^*\|_2 \leq r$, and the bound (9) holds at iteration s . The same argument then implies that the bound (9) also holds

for iteration $s + 1$, and that $\|\theta^{s+1} - \theta^*\|_2 \leq r$. Therefore, by induction the claim holds for each iteration $s \in \{0, 1, 2, 3, \dots\}$.

Using this claim, we can show that

$$\begin{aligned}\|\theta^t - \theta^*\|_2 &\leq \kappa \|\theta^{t-1} - \theta^*\|_2 + \alpha \varepsilon_Q^{\text{unif}}(n, \delta) \\ &\leq \kappa \left\{ \kappa \|\theta^{t-2} - \theta^*\|_2 + \alpha \varepsilon_Q^{\text{unif}}(n, \delta) \right\} + \alpha \varepsilon_Q^{\text{unif}}(n, \delta) \\ &\leq \kappa^t \|\theta^0 - \theta^*\|_2 + \left\{ \sum_{s=0}^{t-1} \kappa^s \right\} \alpha \varepsilon_Q^{\text{unif}}(n, \delta) \\ &\leq \kappa^t \|\theta^0 - \theta^*\|_2 + \frac{\alpha}{1 - \kappa} \varepsilon_Q^{\text{unif}}(n, \delta)\end{aligned}$$

This yields the result in theorem. \square

A.3 Additional Details for the Extension of Results to the EM Algorithm

In this subsection, we develop some results for the extension of first order EM algorithm.

A.3.1 Analysis of the EM Algorithm at the Population Level

We assume throughout this section that the function q is λ -strongly concave (but not necessarily smooth). In order to compactly represent the EM update, we define the operator $M : \Omega \rightarrow \Omega$

$$M(\theta) = \arg \max_{\theta' \in \Omega} Q(\theta' | \theta) \quad (10)$$

Using this notation, the EM algorithm given some initialization θ^0 , produces a sequence of iterates $\{\theta^t\}_{t=0}^\infty$, where $\theta^{t+1} = M(\theta^t)$. By virtue of the self-consistency property and the convexity of Ω , the fixed point satisfies the first-order optimality (KKT) condition

$$\langle \nabla Q(\theta^* | \theta^*), \theta' - \theta^* \rangle \leq 0 \quad \text{for all } \theta' \in \Omega \quad (11)$$

Similarly, for any $\theta \in \Omega$, since $M(\theta)$ maximizes the function $\theta' \mapsto Q(\theta' | \theta)$ over Ω , we have

$$\langle \nabla Q(M(\theta) | \theta), \theta' - M(\theta) \rangle \leq 0 \quad \text{for all } \theta' \in \Omega \quad (12)$$

Now we introduce the following regularity condition in order to relate conditions (12) and (11): The condition involves a Euclidean ball of radius r around the fixed point θ^* , given by

$$\mathbb{B}_2(r; \theta^*) := \{\theta \in \Omega \mid \|\theta - \theta^*\|_2 \leq r\} \quad (13)$$

Definition A.1. The functions $\{Q(\cdot | \theta), \theta \in \Omega\}$ satisfy condition FOS(γ) over $\mathbb{B}_2(r; \theta^*)$ if

$$\|\nabla Q(M(\theta) | \theta^*) - \nabla Q(M(\theta) | \theta)\|_2 \leq \gamma \|\theta - \theta^*\|_2 \quad (14)$$

for all $\theta \in \mathbb{B}_2(r; \theta^*)$.

Theorem 1. For some radius $r > 0$ and pair (γ, λ) such that $0 \leq \gamma < \lambda$ suppose that the function $Q(\cdot | \theta^*)$ is globally λ -strongly concave (5), and that the FOS(γ) condition (14) holds on the ball $\mathbb{B}_2(r; \theta^*)$. Then the population EM operator M is contractive over $\mathbb{B}_2(r; \theta^*)$, in particular with

$$\|M(\theta) - \theta^*\|_2 \leq \frac{\gamma}{\lambda} \|\theta - \theta^*\|_2$$

for all $\theta \in \mathbb{B}_2(r; \theta^*)$.

Proof. Since both $M(\theta)$ and θ^* are in Ω , we may apply condition (11) with $\theta' = M(\theta)$ and condition (12) with $\theta' = \theta^*$ with some algebras. Then we will obtain

$$\langle \nabla Q(M(\theta)|\theta^*) - \nabla Q(\theta^*|\theta^*), \theta^* - M(\theta) \rangle \leq \langle \nabla Q(M(\theta)|\theta^*) - \nabla Q(M(\theta)|\theta), \theta^* - M(\theta) \rangle \quad (15)$$

Applying λ -strong concavity condition (5) on LHS gives us

$$\langle \nabla Q(M(\theta)|\theta^*) - \nabla Q(\theta^*|\theta^*), \theta^* - M(\theta) \rangle \geq \lambda \|\theta^* - M(\theta)\|_2^2 \quad (16)$$

Applying the FOS(γ) condition together with the Cauchy-Schwarz inequality on RHS gives us

$$\langle \nabla Q(M(\theta)|\theta^*) - \nabla Q(M(\theta)|\theta), \theta^* - M(\theta) \rangle \leq \gamma \|\theta^* - M(\theta)\|_2 \|\theta - \theta^*\|_2 \quad (17)$$

Combining the inequalities above the gives us

$$\lambda \|\theta^* - M(\theta)\|_2^2 \leq \gamma \|\theta^* - M(\theta)\|_2 \|\theta - \theta^*\|_2$$

canceling term $\|\theta^* - M(\theta)\|_2$ yields the result. \square

A.3.2 Finite-sample Analysis for the EM Algorithm

We now turn to theoretical results on the sample-based version of the EM algorithm. Define the sample-based operator $M_n : \Omega \rightarrow \Omega$

$$M_n(\theta) = \arg \max_{\theta' \in \Omega} Q_n(\theta'|\theta) \quad (18)$$

where the sample-based Q -function was defined in equation (1). Given the sample size n and T iterations, the full data set is divided into T subsets of size $\lfloor n/T \rfloor$. The mini-batch EM Algorithm updates $\theta^{t+1} = M_{n/T}(\theta^t)$. For the tolerance parameter $\delta \in (0, 1)$, we let $\varepsilon_M(n, \delta)$ be the smallest scalar such that for any fixed $\theta \in \mathbb{B}_2(r; \theta^*)$

$$\mathbb{P}(\|M_n(\theta) - M(\theta)\|_2 \leq \varepsilon_M(n, \delta)) \geq 1 - \delta \quad (19)$$

On the other hand, in order to analyze the standard sample-based form of EM, define $\varepsilon_M^{\text{unif}}(n, \delta)$ to be the smallest scalar for which

$$\sup_{\theta \in \mathbb{B}_2(r; \theta^*)} \|M_n(\theta) - M(\theta)\|_2 \leq \varepsilon_M^{\text{unif}}(n, \delta) \quad (20)$$

with probability at least $1 - \delta$. With these definitions, we have the following guarantees.

Theorem 2. Suppose that the population EM operator $M : \Omega \rightarrow \Omega$ is contractive with parameter $\kappa \in (0, 1)$ on the ball $\mathbb{B}_2(r; \theta^*)$, and the initial vector θ^0 belongs to $\mathbb{B}_2(r; \theta^*)$

(a) If the sample size n is large enough to ensure that

$$\varepsilon_M^{\text{unif}}(n, \delta) \leq (1 - \kappa)r \quad (21)$$

then the EM iterates $\{\theta^t\}_{t=0}^\infty$ satisfy the bound

$$\|\theta^t - \theta^*\|_2 \leq \kappa^t \|\theta^0 - \theta^*\|_2 + \frac{1}{1 - \kappa} \varepsilon_M^{\text{unif}}(n, \delta) \quad (22)$$

(b) For a given iteration number T , suppose the sample size n is large enough to ensure that

$$\varepsilon_M\left(\frac{n}{T}, \frac{\delta}{T}\right) \leq (1 - \kappa)r \quad (23)$$

Then the sample-splitting EM iterates $\{\theta^t\}_{t=0}^T$ based on $\frac{n}{T}$ samples per round satisfy the bound

$$\|\theta^t - \theta^*\|_2 \leq \kappa^t \|\theta^0 - \theta^*\|_2 + \frac{1}{1 - \kappa} \varepsilon_M\left(\frac{n}{T}, \frac{\delta}{T}\right) \quad (24)$$

Remark A.1. The proof for this Theorem is similar to Theorem 2 and Theorem 3.

B Additional Details for the Applications on Specific Statistical Models

B.1 Gaussian Mixture Models for Population Level and Sample-based Methods

Consider the two-component Gaussian mixture model with balanced weights and isotropic covariances. It can be specified by a density of the form

$$f_\theta(y) = \frac{1}{2} \phi(y; \theta^*, \sigma^2 I_d) + \frac{1}{2} \phi(y; -\theta^*, \sigma^2 I_d) \quad (25)$$

where $\phi(\cdot; \mu, \Sigma)$ denotes the density of a $\mathcal{N}(\mu, \Sigma)$ random vector in \mathbb{R}^d , and we have assumed that the two components are equally weighted. Suppose that the variance σ^2 is known, so that our goal is to estimate the unknown mean vector θ^* . In this example, the hidden variable $Z \in \{0, 1\}$ is an indicator variable for the underlying mixture component, that is,

$$(Y|Z=0) \sim \mathcal{N}(-\theta^*, \sigma^2 I_d) \quad \text{and} \quad (Y|Z=1) \sim \mathcal{N}(\theta^*, \sigma^2 I_d)$$

Suppose that we are given n i.i.d. samples $\{y_i\}_{i=1}^n$ drawn from the mixture density (25). In the most general case (two components weighted by $w_\theta(y)$), the complete data $\{(y_i, z_i)\}_{i=1}^n$ corresponds to the original samples along with the component indicator variables $z_i \in \{0, 1\}$. The sample-based function Q_n takes the form

$$Q_n(\theta'|\theta) = -\frac{1}{2n} \sum_{i=1}^n \left[w_\theta(y_i) \|y_i - \theta'\|_2^2 + (1 - w_\theta(y_i)) \|y_i + \theta'\|_2^2 \right] \quad (26)$$

where

$$w_\theta(y) := e^{-\frac{\|\theta-y\|_2^2}{2\sigma^2}} \left[e^{-\frac{\|\theta-y\|_2^2}{2\sigma^2}} + e^{-\frac{\|\theta+y\|_2^2}{2\sigma^2}} \right]^{-1}$$

- EM updates: This example is especially simple in that each iteration of the EM algorithm has a closed form solution, given by

$$\theta^{t+1} := \arg \max_{\theta' \in \mathbb{R}^d} Q_n(\theta'|\theta^t) = \frac{2}{n} \sum_{i=1}^n w_{\theta^t}(y_i) y_i - \frac{1}{n} \sum_{i=1}^n y_i \quad (27)$$

Iterations of the population EM algorithm are specified analogously

$$\theta^{t+1} = 2\mathbb{E}[w_{\theta^t}(Y)Y] \quad (28)$$

where the empirical expectation has been replaced by expectation under the mixture distribution (25).

- First-order EM updates: On the other hand, the sample-based and population first-order EM operators with step size $\alpha > 0$ are given by

$$\theta^{t+1} = \theta^t + \alpha \left\{ \frac{1}{n} \sum_{i=1}^n (2w_{\theta^t}(y_i) - 1) y_i - \theta^t \right\}, \quad \text{and} \quad \theta^{t+1} = \theta^t + \alpha [2\mathbb{E}[w_{\theta^t}(Y)Y] - \theta^t] \quad (29)$$

Before we step into formal analysis on mixture of Gaussians, we provide some technical results related to the mixture of Gaussians

- For the function $f(t) = \frac{t^2}{\exp(\mu t)}$, we have

$$\sup_{t \in [0, \infty]} f(t) = \frac{4}{(e\mu)^2}, \quad \text{achieved at } t^* = \frac{2}{\mu} \quad (30)$$

and

$$\sup_{t \in [t^*, \infty]} f(t) = f(t^*), \quad \text{for } t^* \geq \frac{2}{\mu} \quad (31)$$

- For the function $g(t) = \frac{1}{(\exp(t) + \exp(-t))^2}$, we have

$$g(t) \leq \frac{1}{4} \quad \text{for all } t \in \mathbb{R} \quad (32)$$

and for any $\mu \geq 0$

$$\sup_{t \in [\mu, \infty]} g(t) \leq \frac{1}{(\exp(\mu) + \exp(-\mu))^2} \leq \exp(-2\mu) \quad (33)$$

- Similarly, for the function $g^2(t) = \frac{1}{(\exp(t) + \exp(-t))^4}$, we have

$$g^2(t) \leq \frac{1}{16} \quad \text{for all } t \in \mathbb{R} \quad (34)$$

and for any $\mu \geq 0$

$$\sup_{t \in [\mu, \infty]} g^2(t) \leq \frac{1}{(\exp(\mu) + \exp(-\mu))^4} \leq \exp(-4\mu) \quad (35)$$

B.1.1 Population Level Analysis

The difficulty of estimating such a mixture model can be characterized by the signal-to-noise ratio $\frac{\|\theta^*\|_2}{\sigma}$, and our analysis requires the SNR to be lower bounded as

$$\frac{\|\theta^*\|_2}{\sigma} > \eta \quad (36)$$

With the signal-to-noise ratio lower bound η defined above, we have the following guarantee.

Corollary B.1 (Population result for the first-order EM algorithm for Gaussian). *Consider a Gaussian mixture model for which the SNR condition (36) holds for a sufficiently large η , and define the radius $r = \frac{\|\theta^*\|_2}{4}$. Then there is a contraction coefficient $\kappa(\eta) \leq e^{-c\eta^2}$ where c is a universal constant such that for any initialization $\theta^0 \in \mathbb{B}_2(r; \theta^*)$, the population first-order EM iterates with step size 1, satisfy the bound*

$$\|\theta^t - \theta^*\|_2 \leq \kappa^t \|\theta^0 - \theta^*\|_2 \quad (37)$$

for all $t = 1, 2, \dots$

To prove the Corollary B.1, we need following lemma:

Lemma B.1. Under the conditions of Corollary B.1, there is a constant $\gamma \in (0, 1)$ with $\gamma \leq \exp(-c_2\eta^2)$ such that

$$\|\mathbb{E}[2\Delta_w(Y)Y]\|_2 \leq \gamma \|\theta - \theta^*\|_2 \quad (38)$$

where $\Delta_w(y) := w_\theta(y) - w_{\theta^*}(y)$.

Proof. Denote $\|\cdot\|_{\text{op}}$ to be the operator norm and $\|\cdot\|_{\text{fro}}$ to be the Frobenius norm. For each $u \in [0, 1]$, define $\theta_u = \theta^* + u\Delta$, where $\Delta := \theta - \theta^*$. Applying Taylor's Theorem to the function $\theta \mapsto w_\theta(Y)$ followed by expectations, gives us

$$\mathbb{E}[Y(w_\theta(Y) - w_{\theta^*}(Y))] = 2 \int_0^1 \mathbb{E} \left[\frac{YY^T}{\sigma^2 \left(\exp \left(-\frac{\langle \theta_u, Y \rangle}{\sigma^2} \right) + \exp \left(\frac{\langle \theta_u, Y \rangle}{\sigma^2} \right) \right)^2} \right] \Delta du$$

and denote

$$\Gamma_u(Y) = \frac{YY^T}{\sigma^2 \left(\exp \left(-\frac{\langle \theta_u, Y \rangle}{\sigma^2} \right) + \exp \left(\frac{\langle \theta_u, Y \rangle}{\sigma^2} \right) \right)^2}$$

For each choice of $u \in [0, 1]$, we have $\Gamma_u(y) = \Gamma_u(-y)$. Note that the distribution of Y is symmetric around zero, we see that $\mathbb{E}[\Gamma_u(Y)] = \mathbb{E}[\Gamma_u(\tilde{Y})]$, where $\tilde{Y} \sim \mathcal{N}(\theta^*, \sigma^2 I)$, and hence that

$$\|\mathbb{E}[(w_\theta(Y) - w_{\theta^*}(Y))Y]\|_2 \leq 2 \sup_{u \in [0, 1]} \left\| \mathbb{E}[\Gamma_u(\tilde{Y})] \right\|_{\text{op}} \|\Delta\|_2 \quad (39)$$

Now we need to bound the value of $\left\| \mathbb{E}[\Gamma_u(\tilde{Y})] \right\|_{\text{op}}$ uniformly over $u \in [0, 1]$.

For an arbitrary fixed $u \in [0, 1]$, let R be an rotating matrix such that $R\theta_u = \|\theta_u\|_2 e_1$, where $e_1 \in \mathbb{R}^d$ is the first canonical basis vector. Define the rotated random vector $V = R\tilde{Y}$ and note that $V \sim \mathcal{N}(R\theta^*, \sigma^2 I)$. Using this transformation, the operator norm of the matrix $\mathbb{E}[\Gamma_u(\tilde{Y})]$ is equal to that of

$$D = \mathbb{E} \left[\frac{VV^T}{\sigma^2 \left(\exp \left(\frac{\langle V, \|\theta_u\|_2 e_1 \rangle}{\sigma^2} \right) + \exp \left(-\frac{\langle V, \|\theta_u\|_2 e_1 \rangle}{\sigma^2} \right) \right)^2} \right]$$

Define

$$\begin{aligned} \alpha_1 &:= \mathbb{E} \left[\frac{V_1^2}{\sigma^2 \left(\exp \left(\frac{\langle V, \|\theta_u\|_2 e_1 \rangle}{\sigma^2} \right) + \exp \left(-\frac{\langle V, \|\theta_u\|_2 e_1 \rangle}{\sigma^2} \right) \right)^2} \right] \\ \alpha_2 &:= \mathbb{E} \left[\frac{V_1}{\sigma^2 \left(\exp \left(\frac{\langle V, \|\theta_u\|_2 e_1 \rangle}{\sigma^2} \right) + \exp \left(-\frac{\langle V, \|\theta_u\|_2 e_1 \rangle}{\sigma^2} \right) \right)^2} \right] \\ \alpha_3 &:= \mathbb{E} \left[\frac{1}{\sigma^2 \left(\exp \left(\frac{\langle V, \|\theta_u\|_2 e_1 \rangle}{\sigma^2} \right) + \exp \left(-\frac{\langle V, \|\theta_u\|_2 e_1 \rangle}{\sigma^2} \right) \right)^2} \right] \end{aligned}$$

and

$$\mu := R\theta^* \quad \nu := [0, \mu_2, \mu_3, \dots, \mu_d]^\top$$

Immediately, we have

$$\begin{aligned}
e_1 e_1^\top &= \text{diag}\{1, 0, \dots, 0\} \\
\nu e_1^\top &= \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ \mu_2 & 0 & 0 & \dots & 0 \\ \mu_3 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mu_d & 0 & 0 & \dots & 0 \end{bmatrix} \\
e_1 \nu^\top &= \begin{bmatrix} 0 & \mu_2 & \mu_3 & \dots & \mu_d \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix} \\
e_1 \nu^\top &= \begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ 0 & \mu_2^2 & \mu_2 \mu_3 & \dots & \mu_2 \mu_d \\ 0 & \mu_3 \mu_2 & \mu_3^2 & \dots & \mu_3 \mu_d \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \mu_d \mu_2 & \mu_d \mu_3 & \dots & \mu_d^2 \end{bmatrix}
\end{aligned}$$

Therefore, we have

$$D = \alpha_1 e_1 e_1^\top + \alpha_2 (\nu e_1^\top + e_1 \nu^\top) + \alpha_3 \nu \nu^\top$$

and

$$\|D\|_{\text{op}} \leq \|D\|_{\text{fro}} \leq \alpha_1 + 2\alpha_2 \|\nu\|_2 + \alpha_3 \|\nu\|_2^2 \leq \alpha_1 + 2\alpha_2 \|\theta^*\|_2 + \alpha_3 \|\theta^*\|_2^2 \quad (40)$$

Define the event, we have following claims bound the values α_1 , α_2 and α_3 whenever $\frac{\|\theta^*\|_2^2}{\sigma^2} \geq \eta^2 \geq 16/3$:

Claim B.1.

$$\alpha_1 \leq \frac{16\sigma^2}{9e^2 \|\theta^*\|_2^2} e^{-\frac{\|\theta^*\|_2^2}{32\sigma^2}} + \frac{\|\theta^*\|_2^2}{16\sigma^2} e^{-\frac{3\|\theta^*\|_2^2}{8\sigma^2}}$$

Note that for α_1 , we have

$$\alpha_1 \leq \mathbb{E} \left[\frac{V_1^2 / \sigma^2}{\exp\left(\frac{2\|\theta_u\|_2 V_1}{\sigma^2}\right)} \right]$$

Conditioning on the event \mathcal{E} and \mathcal{E}^c yields

$$\alpha_1 \leq \mathbb{E} \left[\frac{V_1^2 / \sigma^2}{\exp\left(\frac{2\|\theta_u\|_2 V_1}{\sigma^2}\right)} \middle| \mathcal{E} \right] \mathbb{P}[\mathcal{E}] + \mathbb{E} \left[\frac{V_1^2 / \sigma^2}{\exp\left(\frac{2\|\theta_u\|_2 V_1}{\sigma^2}\right)} \middle| \mathcal{E}^c \right] \mathbb{P}[\mathcal{E}^c]$$

Applying equation (30) and (31) yields that when $\|\theta^*\|_2 \|\theta_u\|_2 \geq 4\sigma^2$

$$\alpha_1 \leq \frac{\sigma^2}{e^2 \|\theta_u\|_2^2} \mathbb{P}[\mathcal{E}] + \frac{\|\theta^*\|_2^2}{16\sigma^2 \exp\left(\frac{\|\theta_u\|_2 \|\theta^*\|_2}{2\sigma^2}\right)}$$

Note that in Corollary B.1, we have the condition $r = \frac{\|\theta^*\|_2}{4}$, therefore, we have

$$\|\theta_u\|_2 = \|\theta^* + u(\theta - \theta^*)\|_2 \geq \|\theta^*\|_2 - \frac{1}{4} \|\theta^*\|_2 = \frac{3}{4} \|\theta^*\|_2 \quad (41)$$

Thus, when $\|\theta^*\|_2^2 \geq 16\sigma^2/3$, we have

$$\alpha_1 \leq \frac{16\sigma^2}{9e^2 \|\theta^*\|_2^2} \mathbb{P}(\mathcal{E}) + \frac{\|\theta^*\|_2^2 \exp\left(-\frac{3\|\theta^*\|_2^2}{8\sigma^2}\right)}{16\sigma^2}$$

Using equation 41, the mean of V_1 is lower bounded as

$$\begin{aligned} \mathbb{E}[V_1] &= \langle R\theta^*, e_1 \rangle = \langle R\theta_u, e_1 \rangle + \langle R(\theta^* - \theta_u), e_1 \rangle \\ &\geq \|\theta_u\|_2 - \|\theta^* - \theta_u\|_2 \\ &\geq \frac{\|\theta^*\|_2}{2} \end{aligned}$$

Combining with by standard Gaussian tail bound

$$\mathbb{P}[\mathcal{E}] \leq \exp\left(-\frac{\|\theta^*\|_2^2}{32\sigma^2}\right) \quad (42)$$

gives us when $\|\theta^*\|_2^2 \geq 16\sigma^2/3$

$$\alpha_1 \leq \frac{16\sigma^2}{9e^2 \|\theta^*\|_2^2} e^{-\frac{\|\theta^*\|_2^2}{32\sigma^2}} + \frac{\|\theta^*\|_2^2}{16\sigma^2} e^{-\frac{3\|\theta^*\|_2^2}{8\sigma^2}}$$

This proves our claim on α_1 . The next claim bounds the value of α_2

Claim B.2.

$$\alpha_2 \leq \frac{2\|\theta^*\|_2}{\sigma^2} \exp\left(-\frac{\|\theta^*\|_2^2}{64\sigma^2}\right)$$

Similar to the case for bounding α_1 , we have

$$\begin{aligned} \alpha_2 &= \mathbb{E}\left[\frac{V_1}{\sigma^2 \left(\exp\left(\frac{\|\theta_u\|_2 V_1}{\sigma^2}\right) + \exp\left(-\frac{\|\theta_u\|_2 V_1}{\sigma^2}\right)\right)^2}\right] \\ &\leq \sqrt{\mathbb{E}\left[\frac{V_1^2}{\sigma^2}\right]} \sqrt{\mathbb{E}\left[\frac{1}{\sigma^2 \left(\exp\left(\frac{\|\theta_u\|_2 V_1}{\sigma^2}\right) + \exp\left(-\frac{\|\theta_u\|_2 V_1}{\sigma^2}\right)\right)^4}\right]} \end{aligned}$$

For the first term on RHS of the inequality for α_2 , we have

$$\mathbb{E}\left[\frac{V_1^2}{\sigma^2}\right] \leq \frac{\|\theta^*\|_2^2}{\sigma^2}$$

For the second term on RHS of the inequality for α_2 , applying the equation (32) gives us

$$\mathbb{E}\left[\frac{1}{\sigma^2 \left(\exp\left(\frac{\|\theta_u\|_2 V_1}{\sigma^2}\right) + \exp\left(-\frac{\|\theta_u\|_2 V_1}{\sigma^2}\right)\right)^4}\right] = \frac{1}{\sigma^2} \mathbb{E}\left[g^2\left(\frac{\|\theta_u\|_2 V_1}{\sigma^2}\right)\right]$$

Conditioning on the event \mathcal{E} and \mathcal{E}^c yields

$$\begin{aligned}\frac{1}{\sigma^2} \mathbb{E} \left[g^2 \left(\frac{\|\theta_u\|_2 V_1}{\sigma^2} \right) \right] &\leq \frac{1}{\sigma^2} \left[\mathbb{E} \left[g^2 \left(\frac{\|\theta_u\|_2 V_1}{\sigma^2} \right) \middle| \mathcal{E} \right] \mathbb{P}[\mathcal{E}] + \mathbb{E} \left[g^2 \left(\frac{\|\theta_u\|_2 V_1}{\sigma^2} \right) \middle| \mathcal{E}^c \right] \mathbb{P}[\mathcal{E}^c] \right] \\ &\leq \frac{1}{\sigma^2} \left[\frac{1}{16} \mathbb{P}[\mathcal{E}] + \exp \left(-\frac{\|\theta^*\|_2 \|\theta_u\|_2}{\sigma^2} \right) \right] \\ &\leq \frac{1}{\sigma^2} \left[\frac{1}{16} \mathbb{P}[\mathcal{E}] + \exp \left(-\frac{3 \|\theta^*\|_2^2}{4\sigma^2} \right) \right]\end{aligned}$$

The second line follows by applying bound (32) to the first term, and the bound (33) with

$$\mu = \frac{\|\theta^*\|_2 \|\theta_u\|_2}{4\sigma^2}$$

to the second term. The last inequality follows from the bound (41). Besides, using the bound (42) on $\mathbb{P}[\mathcal{E}]$ gives us

$$\frac{1}{\sigma^2} \mathbb{E} \left[g^2 \left(\frac{\|\theta_u\|_2 V_1}{\sigma^2} \right) \right] \leq \frac{1}{\sigma^2} \left[\frac{1}{16} \exp \left(-\frac{\|\theta^*\|_2^2}{32\sigma^2} \right) + \exp \left(-\frac{3 \|\theta^*\|_2^2}{4\sigma^2} \right) \right] \leq \frac{2}{\sigma^2} \exp \left(-\frac{\|\theta^*\|_2^2}{32\sigma^2} \right)$$

Combining all inequalities together gives us the bound for α_2 :

$$\alpha_2 \leq \frac{2 \|\theta^*\|_2}{\sigma^2} \exp \left(-\frac{\|\theta^*\|_2^2}{64\sigma^2} \right)$$

Claim B.3.

$$\alpha_3 \leq \frac{2}{\sigma^2} \exp \left(-\frac{\|\theta^*\|_2^2}{32\sigma^2} \right)$$

The derivation for α_3 is quite similar to that of α_2 . Again, we first condition on the event \mathcal{E} and \mathcal{E}^c , then apply the bound (32), (33), (41) and (42).

$$\begin{aligned}\alpha_3 &\leq \frac{1}{\sigma^2} \left[\mathbb{E} \left[g \left(\frac{\|\theta_u\|_2 V_1}{\sigma^2} \right) \middle| \mathcal{E} \right] \mathbb{P}[\mathcal{E}] + \mathbb{E} \left[g \left(\frac{\|\theta_u\|_2 V_1}{\sigma^2} \right) \middle| \mathcal{E}^c \right] \mathbb{P}[\mathcal{E}^c] \right] \\ &\leq \frac{1}{\sigma^2} \left[\frac{1}{4} \mathbb{P}[\mathcal{E}] + \exp \left(-\frac{\|\theta^*\|_2 \|\theta_u\|_2}{4\sigma^2} \right) \right] \\ &\leq \frac{1}{\sigma^2} \left[\frac{1}{4} \mathbb{P}[\mathcal{E}] + \exp \left(-\frac{3 \|\theta^*\|_2^2}{16\sigma^2} \right) \right] \\ &\leq \frac{1}{\sigma^2} \left[\frac{1}{4} \exp \left(-\frac{\|\theta^*\|_2^2}{32\sigma^2} \right) + \exp \left(-\frac{3 \|\theta^*\|_2^2}{16\sigma^2} \right) \right] \\ &\leq \frac{2}{\sigma^2} \exp \left(-\frac{\|\theta^*\|_2^2}{32\sigma^2} \right)\end{aligned}$$

Combining with three claims together, we have

$$\|2\mathbb{E}[(w_\theta(Y) - w_{\theta^*}(Y)) Y]\|_2 \leq c_1 \left(1 + \frac{1}{\eta^2} + \eta^2 \right) e^{-c_2 \eta^2} \|\theta - \theta^*\|_2$$

if $\frac{\|\theta^*\|_2^2}{\sigma^2} \geq \eta^2 \geq 16/3$. Therefore, the upper bound (38) holds when the signal-to-noise ratio is sufficiently large. \square

Using this result, we now can prove the Corollary B.1.

Proof of Corollary B.1. Note that scaling the family of Q functions by a fixed constant does not affect any of our conditions and their consequences. Thus, re-scaling Q functions by constant σ^2 does not effect the final results. In order to apply Theorem 1, we need to verify the λ -concavity (5) and μ -smoothness (6) conditions, and the GS(γ) condition (4) over the ball $\mathbb{B}_2(r; \theta^*)$. In this case, the q -function takes the form

$$q(\theta) = Q(\theta|\theta^*) = -\frac{1}{2}\mathbb{E} [w_{\theta^*}(Y)\|Y - \theta\|_2^2 + (1 - w_{\theta^*}(Y))\|Y + \theta\|_2^2]$$

where the weighting function is given by

$$w_\theta(y) := \frac{\exp(-\|\theta - y\|_2^2/(2\sigma^2))}{\exp(-\|\theta - y\|_2^2/(2\sigma^2)) + \exp(-\|\theta + y\|_2^2/(2\sigma^2))}$$

The q -function is smooth and strongly-concave with parameters 1. Using the Lemma above, we now verify the GS condition (4). By smoothness and strong-concavity, we only need to show

$$\|\mathbb{E}[2\Delta_w(Y)Y]\|_2 < \|\theta - \theta^*\|_2$$

This claim follows immediately from the Lemma above. Thus, the GS condition holds when $\gamma < 1$. The bound on the contraction parameter follows from the fact that $\gamma \leq \exp(-c_2\eta^2)$ and applying Theorem 1 yields Corollary B.1. \square

B.1.2 Finite Sample Analysis

Before we step into finite sample analysis on first-order EM algorithm, we introduce some concepts that we may use during the analysis.

Definition B.1. A random variable X is said to be Rademacher Random Variable if it has following density function

$$f(k) = \frac{1}{2}(\delta(k-1) + \delta(k+1)) = \begin{cases} 1/2 & \text{if } k = -1 \\ 1/2 & \text{if } k = +1 \\ 0 & \text{otherwise} \end{cases}$$

We now analyze the convergence rate for sample-based first-order EM updates for the Gaussian mixture model. In this part, the step size is $\alpha = 1$. Consider the function

$$\varphi(\sigma; \|\theta^*\|_2) := \|\theta^*\|_2 \left(1 + \frac{\|\theta^*\|_2^2}{\sigma^2}\right)$$

and positive universal constants (c, c_1, c_2) .

Corollary B.2 (Sample-based first-order EM guarantees for Gaussian mixture). *In addition to the conditions of Corollary B.1, suppose that the sample size is lower bounded as $n \geq c_1 d \log(1/\delta)$. Then given any initialization $\theta^0 \in \mathbb{B}_2\left(\frac{\|\theta^*\|_2}{4}; \theta^*\right)$, there is a contraction coefficient $\kappa(\eta) \leq e^{-cn^2}$ such that the first order EM iterates $\{\theta^t\}_{t=0}^\infty$ satisfy the bound*

$$\|\theta^t - \theta^*\|_2 \leq \kappa^t \|\theta^0 - \theta^*\|_2 + \frac{c_2}{1-\kappa} \varphi(\sigma; \|\theta^*\|_2) \sqrt{\frac{d}{n} \log(1/\delta)} \quad (43)$$

Proof. By definition, with probability at least $1 - \delta$,

$$\sup_{\theta \in \mathbb{B}_2(r; \theta^*)} \|\nabla Q_n(\theta | \theta) - \nabla Q(\theta | \theta)\|_2 \leq \varepsilon_Q^{\text{unif}}(n, \delta)$$

and first-order EM algorithm updates

$$\theta^{t+1} = \theta^t + \alpha \left\{ \frac{1}{n} \sum_{i=1}^n (2w_{\theta^t}(y_i) - 1) y_i - \theta^t \right\}, \quad \text{and} \quad \theta^{t+1} = \theta^t + \alpha [2\mathbb{E}[w_{\theta^t}(Y)Y] - \theta^t]$$

Define the set

$$\mathbb{A} := \left\{ \theta \in \mathbb{R}^d \mid \|\theta - \theta^*\|_2 \leq \|\theta^*\|_2 / 4 \right\}$$

and random variable

$$Z := \sup_{\theta \in \mathbb{A}} \left\| \alpha \left\{ \frac{1}{n} \sum_{i=1}^n (2w_{\theta}(y_i) - 1) y_i - \theta \right\} - \alpha [2\mathbb{E}[w_{\theta}(Y)Y] - \theta] \right\|_2$$

To prove the corollary, we need to give an upper bound on $\varepsilon_Q^{\text{unif}}(n, \delta)$ that is similar to $\varepsilon_Q^{\text{unif}}(n, \delta) \leq (\lambda - \gamma)r$. Now we make following claim on Z

Claim B.4. With sufficiently large constants c_1, c_2 , for $n \geq c_1 d \log(1/\delta)$, we have

$$Z \leq \frac{c_2 \|\theta^*\|_2 (\|\theta^*\|_2^2 + \sigma^2)}{\sigma^2} \sqrt{\frac{d \log(1/\delta)}{n}}$$

with probability at least $1 - \delta$.

We prove this claim by following steps:

- Step 1: Reduce our goal supremum on set \mathbb{A} to a finite maximum over the sphere \mathbb{S}^d .

For each $u \in \mathbb{R}^d$ with $\|u\|_2 = 1$, define

$$Z_u := \sup_{\theta \in \mathbb{A}} \left\{ \frac{1}{n} \sum_{i=1}^n (2w_{\theta}(y_i) - 1) \langle y_i, u \rangle - \mathbb{E}(2w_{\theta}(Y) - 1) \langle Y, u \rangle \right\}$$

Therefore, for any pair (u, v) we have

$$|Z_u - Z_v| \leq Z \|u - v\|_2$$

Let $\{u^1, \dots, u^M\}$ be a $\frac{1}{2}$ -covering of the sphere $S^d = \{v \in \mathbb{R}^d \mid \|v\|_2 = 1\}$. That is, for any $v \in \mathbb{S}^d$, there exists some index $j \in [M]$ such that $\|v - u^j\|_2 \leq \frac{1}{2}$. Then we can write

$$Z_v \leq Z_{u^j} + |Z_v - Z_{u^j}| \leq \max_{j \in [M]} Z_{u^j} + Z \|v - u^j\|_2$$

Combining all results above gives us

$$Z = \sup_{v \in \mathbb{S}^d} Z_v \leq 2 \max_{j \in [M]} Z_{u^j} \tag{44}$$

- Step 2: Introduce a sequence of Rademacher variables.

Let $\{\varepsilon_i\}_{i=1}^n$ be an i.i.d. sequence of Rademacher variables, for any $\lambda > 0$, one can show that

$$\mathbb{E} \left[e^{\lambda Z_w} \right] \leq \mathbb{E} \left[\exp \left(\frac{2}{n} \sup_{\theta \in \mathbb{A}} \sum_{i=1}^n \varepsilon_i (2w_\theta(y_i) - 1) \langle y_i, u \rangle \right) \right]$$

by using a standard symmetrization result for empirical processes given by Koltchinski et al. Besides, for any d -vectors y, θ and θ' , we have the Lipschitz property

$$|2w_\theta(y) - 2w_{\theta'}(y)| \leq \frac{1}{\sigma^2} |\langle \theta, y \rangle - \langle \theta', y \rangle|$$

Using the results on the Ledoux-Talagrand contraction for Rademacher processes given by Koltchinski et al gives us

$$\mathbb{E} \left[\exp \left(\frac{2}{n} \sup_{\theta \in \mathbb{A}} \sum_{i=1}^n \varepsilon_i (2w_\theta(y_i) - 1) \langle y_i, u \rangle \right) \right] \leq \mathbb{E} \left[\exp \left(\frac{4}{n\sigma^2} \sup_{\theta \in \mathbb{A}} \sum_{i=1}^n \varepsilon_i \langle \theta, y_i \rangle \langle y_i, u \rangle \right) \right]$$

For any $\theta \in \mathbb{A}$, we have

$$\|\theta\|_2 \leq \frac{5}{4} \|\theta^*\|_2$$

therefore,

$$\sup_{\theta \in \mathbb{A}} \frac{1}{n} \sum_{i=1}^n \varepsilon_i \langle \theta, y_i \rangle \langle y_i, u \rangle \leq \frac{5}{4} \|\theta^*\|_2 \left\| \frac{1}{n} \sum_{i=1}^n \varepsilon_i y_i y_i^T \right\|_{\text{op}}$$

- Step 3: Bound the random variable Z_u for a fixed $u \in \mathbb{S}^d$.

Based on the results in previous steps, we have

$$\begin{aligned} \mathbb{E} \left[e^{\lambda Z_u} \right] &\leq \mathbb{E} \left[\exp \left(\frac{10\lambda \|\theta^*\|_2}{\sigma^2} \max_{j \in [M]} \frac{1}{n} \sum_{i=1}^n \varepsilon_i \langle y_i, u^j \rangle^2 \right) \right] \\ &\leq \sum_{j=1}^M \mathbb{E} \left[\exp \left(\frac{10\lambda \|\theta^*\|_2}{\sigma^2} \frac{1}{n} \sum_{i=1}^n \varepsilon_i \langle y_i, u^j \rangle^2 \right) \right] \end{aligned} \tag{45}$$

Using Rademacher sign variable, we can represent our Gaussian mixture model as

$$y = \eta \theta^* + w$$

where η is a Rademacher sign variable, and $w \sim \mathcal{N}(0, \sigma^2 I)$. For any $u \in \mathbb{R}^d$, one can show that

$$\mathbb{E} \left[e^{\langle u, y \rangle} \right] = \mathbb{E} \left[e^{\eta \langle u, \theta^* \rangle} \right] \mathbb{E} \left[e^{\langle u, w \rangle} \right] \leq e^{(\|\theta^*\|_2^2 + \sigma^2)/2}$$

This result shows that if $\{y_i\}_{i=1}^n$ are n i.i.d. observations, then they are sub-Gaussians with parameter at most $\gamma = \sqrt{\|\theta^*\|_2^2 + \sigma^2}$. Therefore, $\varepsilon_i \langle y_i, u \rangle^2$ is zero mean sub-exponential, and the MGF is bounded as

$$\mathbb{E} \left[e^{t\varepsilon_i \langle y_i, u \rangle^2} \right] \leq e^{\gamma^4 t^2/2}$$

for all $t > 0$ that is sufficiently small. Combined with the result (45), we will see that

$$\mathbb{E} \left[e^{\lambda Z_u} \right] \leq M e^{c \frac{\lambda^2 \|\theta^*\|_2^2 \gamma^4}{n \sigma^4}} \leq e^{c \frac{\lambda^2 \|\theta^*\|_2^2 \gamma^4}{n \sigma^4} + 2d}$$

Combined with the Chernoff approach, the upper bound on the MGF indicates that for a sufficiently large constant c_1 , when $n \geq c_1 d \log(1/\delta)$ we will see that

$$Z \leq \frac{c_2 \|\theta^*\|_2 \gamma^2}{\sigma^2} \sqrt{\frac{d \log(1/\delta)}{n}}$$

with probability at least $1 - \delta$. This proves our claims and thus yields the Corollary B.2. \square

B.2 Mixture of Regressions (MOR)

Our second example is the mixture of regressions model. In the standard linear regression model, we have i.i.d. observations $(Y, X) \in \mathbb{R} \times \mathbb{R}^d$ and the model given by

$$y_i = \langle x_i, \theta^* \rangle + v_i \quad (46)$$

with following conditions:

- $v_i \sim \mathcal{N}(0, \sigma^2)$ is the independent noise.
- The covariates x_i s are independent with v_i s

Besides, in this example, we also assume that

- $x_i \sim \mathcal{N}(0, I)$
- $\theta^* \in \mathbb{R}^d$ is the unknown

In the mixture of regressions problem, we observe a pair (y_i, x_i) drawn from the model (46) with probability $\frac{1}{2}$, and alternative regression model

$$y_i = \langle x_i, -\theta^* \rangle + v_i$$

with probability $1/2$. Therefore, the latent variables $\{z_i\}_{i=1}^n$ represent the labels of the underlying regression model.

$$z_i = \begin{cases} 1 & \text{if the observation is from (46)} \\ 0 & \text{otherwise} \end{cases}$$

In this problem setting, we derive the EM update rule as follows:

Define the weight function

$$w_\theta(x, y) = \frac{\exp\left(\frac{-(y - \langle x, \theta \rangle)^2}{2\sigma^2}\right)}{\exp\left(\frac{-(y - \langle x, \theta \rangle)^2}{2\sigma^2}\right) + \exp\left(\frac{-(y + \langle x, \theta \rangle)^2}{2\sigma^2}\right)} \quad (47)$$

then the sample-based Q -function is given by

$$Q_n(\theta' | \theta) = -\frac{1}{2n} \sum_{i=1}^n \left(w_\theta(x_i, y_i) (y_i - \langle x_i, \theta' \rangle)^2 + (1 - w_\theta(x_i, y_i)) (y_i + \langle x_i, \theta' \rangle)^2 \right) \quad (48)$$

The sample-based EM algorithm updates θ by maximizing above Q -function. The closed form solution to the optimization problem is given by

$$\theta^{t+1} = \left(\sum_{i=1}^n x_i x_i^T \right)^{-1} \left(\sum_{i=1}^n (2w_{\theta^t}(x_i, y_i) - 1) y_i x_i \right) \quad (49)$$

Similarly, the update rule for population level EM algorithm is given by

$$\theta^{t+1} = 2\mathbb{E}[w_{\theta^t}(X, Y)YX] \quad (50)$$

where the expectation is taken over the joint distribution of the pair $(Y, X) \in \mathbb{R} \times \mathbb{R}^d$. On the other hand, the update rules for first order EM algorithm, for population level and sample-based, are given by

$$\theta^{t+1} = \theta^t + \alpha \left\{ \frac{1}{n} \sum_{i=1}^n [(2w_{\theta^t}(x_i, y_i) - 1)y_i x_i - x_i x_i^T \theta^t] \right\} \quad (51)$$

and

$$\theta^{t+1} = \theta^t + \alpha \mathbb{E}[2w_{\theta^t}(X, Y)YX - \theta^t] \quad (52)$$

where $\alpha > 0$ is a step size parameter.

B.2.1 Population Level Analysis

There are some elementary results on Gaussian random vectors that are useful in proving the main results in this section.

Lemma B.2. Given a Gaussian random vector $X \sim \mathcal{N}(0, I)$ and any fixed vectors $u, v \in \mathbb{R}^d$, then we have

$$\mathbb{E}[\langle X, u \rangle^2 \langle X, v \rangle^2] \leq 3\|u\|_2^2 \|v\|_2^2 \quad (53)$$

and

$$\mathbb{E}[\langle X, u \rangle^4 \langle X, v \rangle^2] \leq 15\|u\|_2^4 \|v\|_2^2 \quad (54)$$

Proof. For any fixed orthonormal matrix $R \in \mathbb{R}^{d \times d}$, $R^T X$ also has a $\mathcal{N}(0, I)$ distribution and

$$\mathbb{E}[\langle X, u \rangle^2 \langle X, v \rangle^2] = \mathbb{E}[\langle X, Ru \rangle^2 \langle X, Rv \rangle^2]$$

By choosing R such that $Ru = \|u\|_2 e_1$ and defining $z = Rv$, we have

$$\begin{aligned} \mathbb{E}[\langle X, Ru \rangle^2 \langle X, Rv \rangle^2] &= \mathbb{E}\left[\|u\|_2^2 X_1^2 \sum_{i=1}^d \sum_{j=1}^d X_i X_j z_i z_j\right] \\ &= \|u\|_2^2 (3z_1^2 + (\|z\|_2^2 - z_1^2)) \\ &\leq 3\|u\|_2^2 \|z\|_2^2 \\ &= 3\|u\|_2^2 \|v\|_2^2 \end{aligned}$$

A similar argument yields the second claim. \square

As in our analysis of the Gaussian mixture model, our theory applies when the signal-to-noise ratio is sufficiently large, as enforced by a condition of the form

$$\frac{\|\theta^*\|_2}{\sigma} > \eta \quad (55)$$

for a sufficiently large constant $\eta > 0$. Under a suitable lower bound on this quantity, our first result guarantees that the first-order EM algorithm is locally convergent to the global optimum θ^* and provides a quantification of the local region of convergence.

Corollary B.3 (Population result for the first-order EM algorithm for MOR). *Consider any mixture of regressions model satisfying the SNR condition (55) for a sufficiently large constant η , and define the radius $r := \frac{\|\theta^*\|_2}{32}$. Then for any $\theta^0 \in \mathbb{B}_2(r; \theta^*)$, the population first-order EM iterates with stepsize 1, satisfy the bound*

$$\|\theta^t - \theta^*\|_2 \leq \left(\frac{1}{2}\right)^t \|\theta^0 - \theta^*\|_2 \quad (56)$$

for $t = 1, 2, \dots$

Before we prove the Corollary B.3, we need following lemma.

Lemma B.3. Under the conditions of Corollary B.3, there is a constant $\gamma < 1/4$ such that for any fixed vector $\tilde{\Delta}$ we have

$$\left| \mathbb{E} \left[\Delta_w(X, Y)(2Z - 1) \langle X, \theta^* \rangle \langle X, \tilde{\Delta} \rangle \right] \right| \leq \frac{\gamma}{2} \|\Delta\|_2 \|\tilde{\Delta}\|_2 \quad (57)$$

and

$$\left| \mathbb{E} \left[\Delta_w(X, Y)v \langle X, \tilde{\Delta} \rangle \right] \right| \leq \frac{\gamma}{2} \|\Delta\|_2 \|\tilde{\Delta}\|_2 \quad (58)$$

Proof. Since the standard deviation σ is known, we can assume that $\sigma = 1$ by a simple re-scaling, then the weight function in (47) is given by

$$w_\theta(x, y) = \frac{\exp\left(\frac{-(y - \langle x, \theta \rangle)^2}{2}\right)}{\exp\left(\frac{-(y - \langle x, \theta \rangle)^2}{2}\right) + \exp\left(\frac{-(y + \langle x, \theta \rangle)^2}{2}\right)} \quad (59)$$

Note that $\Delta = \theta - \theta^*$ and $\tilde{\Delta}$ is any fixed vector in $\mathbb{R}^d \setminus \{0\}$. We can define $\theta_u = \theta^* + u\Delta$ for a scalar $u \in [0, 1]$ and our assumptions guarantee that

$$\|\Delta\|_2 \leq \frac{\|\theta^*\|_2}{32}, \quad \text{and} \quad \|\theta^*\|_2 \geq \eta \quad (60)$$

and we also have

$$\|\theta_u\|_2 \geq \|\theta^*\|_2 - \|\Delta\|_2 \geq \frac{\|\theta^*\|_2}{2} \quad (61)$$

- Proof of Inequality (57). We split the proof of this bound into two separate cases

- Case-1: $\|\Delta\|_2 \leq 1$

In this case, we have

$$\frac{d}{d\theta} w_\theta(X, Y) = \frac{2YX}{(\exp(Y\langle X, \theta \rangle) + \exp(-Y\langle X, \theta \rangle))^2}$$

Taylor series with integral form remainder on the function $\theta \mapsto w_\theta(X, Y)$ gives us

$$\Delta_w(X, Y) = \int_0^1 \frac{2Y\langle X, \Delta \rangle}{(\exp(Z_u) + \exp(-Z_u))^2} du \quad (62)$$

where $Z_u := Y\langle X, \theta^* + u\Delta \rangle$. Substituting for $\Delta_w(X, Y)$ in inequality (57), we see that it suffices to show

$$A_u := \int_0^1 \mathbb{E} \left[\frac{2Y\langle X, \theta^* \rangle}{(\exp(Z_u) + \exp(-Z_u))^2} (2Z - 1)\langle X, \Delta \rangle \langle X, \tilde{\Delta} \rangle \right] du \leq \frac{\gamma}{2} \|\Delta\|_2 \|\tilde{\Delta}\|_2 \quad (63)$$

for some $\gamma \in [0, 1/4)$. To show this result, we need following auxiliary results.

Claim B.5. There is $a\gamma \in [0, 1/4)$ such that for each $u \in [0, 1]$, we have

$$\sqrt{\mathbb{E} \left[\frac{Y^2 \langle X, \theta_u \rangle^2}{(\exp(Z_u) + \exp(-Z_u))^4} \right]} \leq \frac{\gamma}{14} \quad (64)$$

and

$$\sqrt{\mathbb{E} \left[\frac{Y^2}{(\exp(Z_u) + \exp(-Z_u))^4} \right]} \leq \frac{\gamma}{32} \quad (65)$$

whenever $\|\Delta\|_2 \leq 1$.

Using this claim, we can bound the quantity A_u from equation (63). Note that $\theta^* = \theta_u - u\Delta$, we have $A_u = B_1 + B_2$, where

$$\begin{aligned} B_1 &:= \mathbb{E} \left[\frac{2Y \langle X, \theta_u \rangle}{(\exp(Z_u) + \exp(-Z_u))^2} (2Z - 1) \langle X, \Delta \rangle \langle X, \tilde{\Delta} \rangle \right] \\ B_2 &:= -\mathbb{E} \left[\frac{2Yu \langle X, \Delta \rangle}{(\exp(Z_u) + \exp(-Z_u))^2} (2Z - 1) \langle X, \Delta \rangle \langle X, \tilde{\Delta} \rangle \right] \end{aligned}$$

To show $A_u \leq \frac{\gamma}{2} \|\Delta\|_2 \|\tilde{\Delta}\|_2$, we need prove that $\max \{B_1, B_2\} \leq \frac{\gamma}{4} \|\Delta\|_2 \|\tilde{\Delta}\|_2$. Thus, it is natural to bound B_1 and B_2 .

* Bounding B_1 : By the Cauchy-Schwarz inequality, we have

$$\begin{aligned} B_1 &\leq \sqrt{\mathbb{E} \left[\frac{y^2 \langle X, \theta_u \rangle^2}{(\exp(Z_u) + \exp(-Z_u))^4} \right]} \sqrt{\mathbb{E} \left[4(2Z - 1)^2 \langle X, \Delta \rangle^2 \langle X, \tilde{\Delta} \rangle^2 \right]} \\ &\leq \frac{\gamma}{14} \sqrt{\mathbb{E} \left[4 \langle X, \Delta \rangle^2 \langle X, \tilde{\Delta} \rangle^2 \right]} \end{aligned}$$

where the second step follows from the bound (64) in the claim above, and the fact that $(2Z - 1)^2 = 1$.

Next we observe that $\mathbb{E} \left[4 \langle X, \Delta \rangle^2 \langle X, \tilde{\Delta} \rangle^2 \right] \leq 12 \|\Delta\|_2^2 \|\tilde{\Delta}\|_2^2$, where we have used the bound (53) in Lemma B.2. Combined with our earlier bound, we have

$$B_1 \leq \frac{\gamma}{4} \|\Delta\|_2 \|\tilde{\Delta}\|_2$$

as claimed.

* Bounding B_2 : Similarly, by Cauchy-Schwarz inequality we have

$$\begin{aligned} B_2 &\leq \sqrt{\mathbb{E} \left[\frac{y^2}{(\exp(Z_u) + \exp(-Z_u))^4} \right]} \sqrt{\mathbb{E} \left[4u^2 (2Z - 1)^2 \langle X, \Delta \rangle^4 \langle X, \tilde{\Delta} \rangle^2 \right]} \\ &\leq \frac{\gamma}{32} \sqrt{\mathbb{E} \left[4u^2 \langle X, \Delta \rangle^4 \langle X, \tilde{\Delta} \rangle^2 \right]} \end{aligned}$$

where the second step follows from the bound (65), and the fact that $(2Z - 1)^2 = 1$. In this case, we have

$$\mathbb{E} \left[4u^2 \langle X, \Delta \rangle^4 \langle X, \tilde{\Delta} \rangle^2 \right] \leq 60 \|\Delta\|_2^4 \|\tilde{\Delta}\|_2^2 \leq 60 \|\Delta\|_2^2 \|\tilde{\Delta}\|_2^2$$

where the first step uses the bound (54) from Lemma B.2, and the second step follows from $\|\Delta\|_2 \leq 1$. Combining these two results, we have

$$B_2 \leq \frac{\gamma}{4} \|\Delta\|_2 \|\tilde{\Delta}\|_2$$

which completes the proof of inequality (57) in the case $\|\Delta\|_2 \leq 1$

- Case-2: $\|\Delta\|_2 \geq 1$.

Our argument in this case makes use of various probability bounds on different events, which we state here for future reference. These events are related to the scalar $\tau := C_\tau \sqrt{\log \|\theta^*\|_2}$ for a constant C_τ , and are also related to the vectors

$$\Delta := \theta - \theta^*, \text{ and } \theta_u := \theta^* + u\Delta \text{ for some fixed } u \in [0, 1]$$

Claim B.6 (Event bounds). We have following probability bounds for the events given below

- (i) For the event $\mathcal{E}_1 := \{\text{sign}(\langle X, \theta^* \rangle) = \text{sign}(\langle X, \theta_u \rangle)\}$, we have $\mathbb{P}[\mathcal{E}_1^c] \leq \frac{\|\Delta\|_2}{\|\theta^*\|_2}$
- (ii) For the event $\mathcal{E}_2 := \{|\langle X, \theta^* \rangle| > \tau\} \cap \{|\langle X, \theta_u \rangle| > \tau\} \cap \{|v| \leq \frac{\tau}{2}\}$, we have

$$\mathbb{P}[\mathcal{E}_2^c] \leq \frac{\tau}{\|\theta^*\|_2} + \frac{\tau}{\|\theta_u\|_2} + 2 \exp\left(-\frac{\tau^2}{2}\right)$$

- (iii) For the event $\mathcal{E}_3 := \{|\langle X, \theta^* \rangle| \geq \tau\} \cup \{|\langle X, \theta_u \rangle| \geq \tau\}$, we have $\mathbb{P}[\mathcal{E}_3^c] \leq \frac{\tau}{\|\theta^*\|_2} + \frac{\tau}{\|\theta_u\|_2}$
- (iv) For the event $\mathcal{E}_4 := \{|v| \leq \tau/2\}$, we have $\mathbb{P}[\mathcal{E}_4^c] \leq 2e^{-\frac{\tau^2}{2}}$
- (v) For the event $\mathcal{E}_5 := \{|\langle X, \theta_u \rangle| > \tau\}$, we have $\mathbb{P}[\mathcal{E}_5^c] \leq \frac{\tau}{\|\theta_u\|_2}$
- (vi) For the event $\mathcal{E}_6 := \{|\langle X, \theta^* \rangle| > \tau\}$, we have $\mathbb{P}[\mathcal{E}_6^c] \leq \frac{\tau}{\|\theta^{-1}\|_2}$

Besides, we will also use the result on controlling the second moment matrix $\mathbb{E}[XX^T]$ when conditioned on some of the events given above:

Claim B.7 (Conditional covariance bounds). Conditioned on any event $\mathcal{E} \in \{\mathcal{E}_1 \cap \mathcal{E}_2, \mathcal{E}_1^c, \mathcal{E}_5^c, \mathcal{E}_6^c\}$, we have

$$\|\mathbb{E}[XX^T | \mathcal{E}]\|_{op} \leq 2$$

Here our goal is to bound the quantity

$$T = \left| \mathbb{E} \left[\Delta_w(X, Y)(2Z - 1) \langle X, \theta^* \rangle \langle X, \tilde{\Delta} \rangle \right] \right| \leq \mathbb{E} \left[\left| \Delta_w(X, Y)(2Z - 1) \langle X, \theta^* \rangle \langle X, \tilde{\Delta} \rangle \right| \right]$$

For any measurable event \mathcal{E} , we define

$$\Psi(\mathcal{E}) := \mathbb{E} \left[\left| \Delta_w(X, Y)(2Z - 1) \langle X, \theta^* \rangle \langle X, \tilde{\Delta} \rangle \right| \middle| \mathcal{E} \right] \mathbb{P}[\mathcal{E}]$$

and note that by successive conditioning, we have

$$T \leq \Psi(\mathcal{E}_1 \cap \mathcal{E}_2) + \Psi(\mathcal{E}_1^c) + \Psi(\mathcal{E}_4^c) + \Psi(\mathcal{E}_5^c) + \Psi(\mathcal{E}_6^c) \quad (66)$$

We bound each of these five terms in turn and summarized the results in following claim.

Claim B.8.

$$\begin{aligned}
\Psi(\mathcal{E}_1 \cap \mathcal{E}_2) &\leq 2\|\tilde{\Delta}\|_2 \|\theta^*\|_2 e^{-\tau^2} \\
\Psi(\mathcal{E}_1^c) &\leq \sqrt{\mathbb{E}[(X, \tilde{\Delta})^2 | \mathcal{E}_1^c]} \sqrt{\mathbb{E}[(X, \Delta)^2 | \mathcal{E}_1^c]} \frac{\|\Delta\|_2}{\|\theta^*\|_2} \leq \frac{2\|\tilde{\Delta}\|_2 \|\Delta\|_2^2}{\|\theta^*\|_2} \\
\Psi(\mathcal{E}_5^c) &\leq \frac{2\tau\|\tilde{\Delta}\|_2 \sqrt{\tau^2 + 2\|\Delta\|_2^2}}{\|\theta_u\|_2} \leq \frac{2\tau\|\tilde{\Delta}\|_2 \|\Delta\|_2 \sqrt{\tau^2 + 2}}{\|\theta_u\|_2} \\
\Psi(\mathcal{E}_6^c) &\leq \frac{\sqrt{2}\tau^2\|\tilde{\Delta}\|_2}{\|\theta^*\|_2}
\end{aligned}$$

We have thus obtained bounds on all five terms in the decomposition (66). We combine these bounds with the lower bound $\|\theta_u\|_2 \geq \frac{\|\theta^*\|_2}{2}$ from equation (61) and then perform some algebra to obtain

$$T \leq c\|\Delta\|_2\|\tilde{\Delta}\|_2 \left\{ \frac{\tau^2}{\|\theta^*\|_2} + \|\theta^*\|_2 e^{-\tau^2/2} \right\} + 2\|\tilde{\Delta}\|_2 \frac{\|\Delta\|_2^2}{\|\theta^*\|_2}$$

where c is a universal constant. In particular, selecting $\tau = c_\tau \sqrt{\log \|\theta^*\|_2}$ for a sufficient large constant c_τ and the constant η in (60) sufficiently large yields the inequality (57).

- Proof of Inequality (58): Similarly, we split the proof of this bound into two separate cases

- Case-1: $\|\Delta\|_2 \leq 1$

As before, by using Taylor expansion of the function $\theta \mapsto \Delta_w(X, Y)$, we only need to prove

$$\int_0^1 \mathbb{E} \left[\frac{2Yv}{(\exp(Z_u) + \exp(-Z_u))^2} \langle X, \Delta \rangle \langle X, \tilde{\Delta} \rangle \right] du \leq \frac{\gamma}{2} \|\Delta\|_2 \|\tilde{\Delta}\|_2$$

For any fixed $u \in [0, 1]$, we have

$$\begin{aligned}
\mathbb{E} \left[\frac{2Yv \langle X, \Delta \rangle \langle X, \tilde{\Delta} \rangle}{(\exp(Z_u) + \exp(-Z_u))^2} \right] &\leq \sqrt{\mathbb{E} \left[\frac{4Y^2}{(\exp(Z_u) + \exp(-Z_u))^4} \right]} \sqrt{\mathbb{E} [v^2 \langle X, \Delta \rangle^2 \langle X, \tilde{\Delta} \rangle^2]} \\
&\leq \sqrt{\mathbb{E} \left[\frac{4Y^2}{(\exp(Z_u) + \exp(-Z_u))^4} \right]} \sqrt{3\|\Delta\|_2^2 \|\tilde{\Delta}\|_2^2} \\
&\leq \frac{\sqrt{3}\gamma}{16} \|\Delta\|_2 \|\tilde{\Delta}\|_2
\end{aligned}$$

The first step follows from Cauchy-Schwarz inequality. The second step follows from inequality (53) in Lemma B.2, the independence of v and X , and the fact that $\mathbb{E}[v^2] = 1$. The last step follows from the bound (65) in claim B.5.

- Case-2: $\|\Delta\|_2 \geq 1$

By Cauchy-Schwarz inequality, it suffices show that

$$\sqrt{\mathbb{E} [\Delta_w^2(X, Y)]} \leq \frac{\gamma}{2}$$

In claim B.6, we have introduce the scalar $\tau := C_\tau \sqrt{\log \|\theta^*\|_2}$ and the events \mathcal{E}_1 and \mathcal{E}_2 . For any measurable event \mathcal{E} , define the function

$$\Psi(\mathcal{E}) = \mathbb{E} [\Delta_w^2(X, Y) | \mathcal{E}] \mathbb{P}[\mathcal{E}]$$

With this notation, by successive conditioning, we have the upper bound

$$\mathbb{E} [\Delta_w^2(X, Y)] \leq \Psi(\mathcal{E}_1^c) + \Psi(\mathcal{E}_1 \cap \mathcal{E}_2^c) + \Psi(\mathcal{E}_1 \cap \mathcal{E}_2) \quad (67)$$

We control each of these terms in turn and summarize the results in following claim

Claim B.9.

$$\begin{aligned} \Psi(\mathcal{E}_1^c) &\leq 4\mathbb{P}[\mathcal{E}_1^c] \leq 4 \frac{\|\Delta\|_2}{\|\theta^*\|_2} \\ \Psi(\mathcal{E}_1 \cap \mathcal{E}_2^c) &\leq 4\mathbb{P}[\mathcal{E}_2^c] \leq 4 \left\{ \frac{\tau}{\|\theta^*\|_2} + \frac{\tau}{\|\theta_u\|_2} + 2e^{-\frac{\tau^2}{2}} \right\} \\ \Psi(\mathcal{E}_1 \cap \mathcal{E}_2) &\leq e^{-2\tau^2} \end{aligned}$$

Now putting them together yields

$$\sqrt{\mathbb{E}[\Delta_w^2(X, Y)]} \leq \sqrt{4 \frac{\|\Delta\|_2}{\|\theta^*\|_2} + 4 \left\{ \frac{\tau}{\|\theta^*\|_2} + \frac{\tau}{\|\theta_u\|_2} + 2e^{-\frac{\tau^2}{2}} \right\} + e^{-2\tau^2}}$$

By choosing C_τ sufficiently large in the definition of τ and selecting the signal-to-noise constant η in condition (60) sufficiently large, the inequality follows. \square

Remark B.1. For detailed proof of Claim B.5, Claim B.6, Claim B.2.1 and Claim B.9, you can refer to Statistical Guarantees for the EM Algorithm: From Population to Sample-based Analysis

With above Lemmas established above, we now can prove the Corollary B.3.

Proof of Corollary B.3. To prove the corollary, we need to verify the condition (3.2) (λ -strong concavity) condition (3.3) (μ -smoothness) and condition (3.1) the GS(γ) over the ball $\mathbb{B}_2(r; \theta^*)$. Under the framework of MOR, the q -function is

$$q(\theta) = Q(\theta | \theta^*) = -\frac{1}{2}\mathbb{E} [w_{\theta^*}(X, Y)(Y - \langle X, \theta \rangle)^2 + (1 - w_{\theta^*}(X, Y))(Y + \langle X, \theta \rangle)^2]$$

where

$$w_\theta(x, y) := \frac{\exp(-(y - \langle x, \theta \rangle)^2 / (2\sigma^2))}{\exp(-(y - \langle x, \theta \rangle)^2 / (2\sigma^2)) + \exp(-(y + \langle x, \theta \rangle)^2 / (2\sigma^2))}$$

Note that function $Q(\cdot | \theta^*)$ is λ -strongly concave and μ -smooth with λ and μ equal to the smallest and largest (resp.) eigenvalue of the matrix $\mathbb{E}[XX^T]$, thus the strong concavity and smoothness hold with $\lambda = \mu = 1$ since $\mathbb{E}[XX^T] = I$ by assumption.

To verify condition GS, we define the difference function

$$\Delta_w(X, Y) := w_\theta(X, Y) - w_{\theta^*}(X, Y)$$

and the difference vector

$$\Delta = \theta - \theta^*$$

Using the updates given by equation (52), we need to show that

$$\|2\mathbb{E}[\Delta_w(X, Y)YX]\|_2 < \|\Delta\|_2$$

This is equivalent to

$$\left\langle 2\mathbb{E}[\Delta_w(X, Y)YX], \tilde{\Delta} \right\rangle < \|\Delta\|_2 \|\tilde{\Delta}\|_2 \quad \text{for } \tilde{\Delta} \in \mathbb{R}^d \setminus \{0\}$$

Under the framework of MOR, $Y \stackrel{d}{=} (2Z - 1)\langle X, \theta^* \rangle + v$ where $Z \sim \text{Bernoulli}(1/2)$ and $v \sim \mathcal{N}(0, 1)$. Thus, it suffices to show that

$$\mathbb{E}[\Delta_w(X, Y)(2Z - 1)\langle X, \theta^* \rangle \langle X, \tilde{\Delta} \rangle] + \mathbb{E}[\Delta_w(X, Y)v\langle X, \tilde{\Delta} \rangle] \leq \gamma \|\Delta\|_2 \|\tilde{\Delta}\|_2 \quad (68)$$

for $\gamma \in [0, 1/2]$ to establish contractivity. In order to prove the theorem with the desired upper bound on the coefficient of contraction we need to show (68) with $\gamma \in [0, 1/4]$. In Lemma, two bounds in conjunction imply that

$$\left\langle \mathbb{E}[\Delta_w(X, Y)YX], \tilde{\Delta} \right\rangle \leq \gamma \|\Delta\|_2 \|\tilde{\Delta}\|_2$$

with $\gamma \in [0, 1/4]$ and thus the corollary B.3 holds. \square

Remark B.2. Note that in Gaussian mixture model the population likelihood has global maxima at θ^* and $-\theta^*$, and a local minimum at 0. Therefore, the radius of Euclidean ball over which the iterates could converge to θ^* should be less than $\|\theta^*\|_2$.

B.2.2 Sample-based Analysis

Our result on sample based analysis for MOR is related to the quantity

$$\varphi(\sigma; \|\theta^*\|_2) = \sqrt{\sigma^2 + \|\theta^*\|_2^2}$$

with positive universal constants (c_1, c_2)

Corollary B.4. *In addition to the conditions of Corollary B.3, suppose that the sample size is lower bounded as $n \geq c_1 d \log(T/\delta)$. Then there is a contraction coefficient $\kappa \leq 1/2$ such that, for any initial vector $\theta^0 \in \mathbb{B}_2\left(\frac{\|\theta^*\|_2}{32}; \theta^*\right)$, the sample-splitting first-order EM iterates with stepsize 1, based on n/T samples per step satisfy the bound*

$$\|\theta^t - \theta^*\|_2 \leq \kappa^t \|\theta^0 - \theta^*\|_2 + c_2 \varphi(\sigma; \|\theta^*\|_2) \sqrt{\frac{d}{n} T \log(T/\delta)} \quad (69)$$

with probability at least $1 - \delta$

Proof. Similar to Corollary B.2, our goal is to bound $\varepsilon_Q(n, \delta)$ defined as

$$\mathbb{P}\left[\left\|\nabla Q_n(\theta|\theta^t)|_{\theta=\theta^t} - \nabla Q(\theta|\theta^t)|_{\theta=\theta^t}\right\|_2 > \varepsilon_Q(n, \delta)\right] \leq 1 - \delta$$

For the first-order EM updates for MOR, we need to control the random variable,

$$Z := \left\| \alpha \left\{ \frac{1}{n} \sum_{i=1}^n (2w_\theta(y_i) - 1) y_i - \theta \right\} - \alpha [2\mathbb{E}[w_\theta(Y)Y] - \theta] \right\|_2$$

Claim B.10. There are universal constants (c_1, c_2) such that given a sample size $n \geq c_1 d \log(1/\delta)$, we have

$$\mathbb{P}\left[Z > \frac{c_2 \|\theta^*\|_2 (\|\theta^*\|_2^2 + \sigma^2)}{\sigma^2} \sqrt{\frac{d \log(1/\delta)}{n}}\right] \leq \delta$$

When step size is $\alpha = 1$, we have

$$Z \leq \left\| \frac{1}{n} \sum_{i=1}^n (2w_\theta(x_i, y_i) - 1) y_i x_i - \mathbb{E}(2w_\theta(X, Y) - 1) Y X \right\|_2 + \left\| I - \frac{1}{n} \sum_{i=1}^n x_i x_i^T \right\|_{\text{op}} \|\theta\|_2$$

Define

$$\begin{aligned} \hat{\Sigma} &:= \frac{1}{n} \sum_{i=1}^n x_i x_i^T \quad \text{with } \Sigma = \mathbb{E}[XX^T] = I \\ \hat{v} &:= \frac{1}{n} \sum_{i=1}^n [\mu_\theta(x_i, y_i) y_i x_i] \quad \text{with } v := \mathbb{E}[\mu_\theta(X, Y) Y X] \end{aligned}$$

where $\mu_\theta(x, y) := 2w_\theta(x, y) - 1$. Noting that $\mathbb{E}[YX] = 0$, we have the bound

$$Z \leq \underbrace{\|\hat{v} - v\|_2}_{T_1} + \underbrace{\|\hat{\Sigma} - \Sigma\|_{\text{op}} \|\theta\|_2}_{T_2} \tag{70}$$

We bound each of the terms T_1 and T_2 in turn.

- Bounding T_1 : Let us write $\|\hat{v} - v\|_2 = \sup_{u \in \mathbb{S}^d} Z(u)$, where

$$Z(u) := \frac{1}{n} \sum_{i=1}^n \mu_\theta(x_i, y_i) y_i \langle x_i, u \rangle - \mathbb{E}[\mu_\theta(X, Y) Y \langle X, u \rangle]$$

The discretization over a $1/2$ -cover of the sphere \mathbb{S}^d – say $\{u^1, \dots, u^M\}$ gives us

$$\|\hat{v} - v\|_2 \leq 2 \max_{j \in [M]} Z(u^j)$$

Thus, it suffices to control the random variable $Z(u)$ for a fixed $u \in \mathbb{S}^d$. By a standard symmetrization argument we have

$$\mathbb{P}[Z(u) \geq t] \leq 2\mathbb{P}\left[\frac{1}{n} \sum_{i=1}^n \varepsilon_i \mu_\theta(x_i, y_i) y_i \langle x_i, u \rangle \geq t/2\right]$$

where $\{\varepsilon_i\}_{i=1}^n$ are an i.i.d. sequence of Rademacher variables. Define

$$\mathcal{E} = \left\{ \frac{1}{n} \sum_{i=1}^n \langle x_i, u \rangle^2 \leq 2 \right\}$$

Then we have $\mathbb{P}[\mathcal{E}^c] \leq e^{-n/32}$ because each variable $\langle x_i, u \rangle$ is sub-Gaussian with parameter one. Therefore, we have

$$\mathbb{P}[Z(u) \geq t] \leq 2\mathbb{P}\left[\frac{1}{n} \sum_{i=1}^n \varepsilon_i \mu_\theta(x_i, y_i) y_i \langle x_i, u \rangle \geq \frac{t}{2} |\mathcal{E}^c\} + 2e^{-n/32}\right]$$

Using the Ledoux-Talagrand contraction for Rademacher processes and the fact that $|\mu_\theta(x, y)| \leq 1$ for all pairs (x, y) gives us

$$\mathbb{E}\left[\exp\left(\frac{\lambda}{n} \sum_{i=1}^n \varepsilon_i \mu_\theta(x_i, y_i) y_i \langle x_i, u \rangle\right) \middle| \mathcal{E}^c\right] \leq \mathbb{E}\left[\exp\left(\frac{2\lambda}{n} \sum_{i=1}^n \varepsilon_i y_i \langle x_i, u \rangle\right) \middle| \mathcal{E}^c\right]$$

Note that conditioned on x_i , y_i is zero-mean and sub-Gaussian with parameter at most $\sqrt{\|\theta^*\|_2^2 + \sigma^2}$. Therefore, we have

$$\begin{aligned}\mathbb{E} \left[\exp \left(\frac{2\lambda}{n} \sum_{i=1}^n \varepsilon_i y_i \langle x_i, u \rangle \right) \middle| \mathcal{E} \right] &\leq \left[\exp \left(\frac{4\lambda^2}{n^2} (\|\theta^*\|_2^2 + \sigma^2) \sum_{i=1}^n \langle x_i, u \rangle^2 \right) \middle| \mathcal{E} \right] \\ &\leq \exp \left(\frac{8\lambda^2}{n} (\|\theta^*\|_2^2 + \sigma^2) \right)\end{aligned}$$

where expectations are taken over the distribution $(y_i|x_i)$ for each index i and the second inequality follows from the definition of \mathcal{E} . Now applying this bound on the MGF gives us

$$\mathbb{P} \left[\frac{1}{n} \sum_{i=1}^n \varepsilon_i \mu_\theta(x_i, y_i) y_i \langle x_i, u \rangle \geq t/2 \middle| \mathcal{E} \right] \leq \exp \left(-\frac{nt^2}{256 (\|\theta^*\|_2^2 + \sigma^2)} \right)$$

Note that $1/2$ -cover of the unit sphere \mathbb{S}^d has at most 2^d elements. Therefore, there is a universal constant c such that

$$T_1 \leq c \sqrt{\|\theta^*\|_2^2 + \sigma^2} \sqrt{\frac{d}{n} \log(1/\delta)}$$

with probability at least $1 - \delta$.

- Bouding T_2 :

On the one hand, we have

$$\|\hat{\Sigma} - \Sigma\|_{\text{op}} \leq c \sqrt{\frac{d}{n} \log(1/\delta)}$$

with probability at least $1 - \delta$ based on standard results in random matrix theory. On the other hand, note that each iteration decreases the distance to θ^* and the iterate satisfies $\|\theta\|_2 \leq 2 \|\theta^*\|_2$. Therefore, we have

$$T_2 \leq c \|\theta^*\|_2 \sqrt{\frac{d}{n} \log(1/\delta)}$$

with probability at least $1 - \delta$.

Finally, applying bounds on T_1 and T_2 on the decomposition (70) yields the claim. \square

Remark B.3. In Corollary B.4, the bound (69) provides guidance on the number of iterations to perform. For a given sample size n , suppose we perform $T = \lceil \log(n/d\varphi^2(\sigma; \|\theta^*\|_2)) \rceil$ iterations. Then based on the bound (69), we have

$$\|\theta^T - \theta^*\|_2 \leq c_3 \varphi(\sigma; \|\theta^*\|_2) \sqrt{\frac{d}{n} \log^2 \left(\frac{n}{d\varphi^2(\sigma; \|\theta^*\|_2)} \right) \log(1/\delta)} \quad (71)$$

with probability at least $1 - \delta$. Besides, Corollary B.4 predicts that the statistical error $\|\theta^t - \theta^*\|_2$ should decrease geometrically and then end at a plateau.

B.3 Linear Regression with Missing Covariates

In traditional linear regression, our training/testing data $(y_i, x_i) \in \mathbb{R} \times \mathbb{R}^d$ are generated based on the linear model (46). Instead of observing the covariate vector $x_i \in \mathbb{R}^d$ directly, we now deal with the case that some of covariates might be missing. Consider $\tilde{x}_i \in \mathbb{R}^d$ with components

$$\tilde{x}_{ij} = \begin{cases} x_{ij}, & \text{with probability } 1 - \rho \\ *, & \text{with probability } \rho \end{cases} \quad (72)$$

where $\rho \in [0, 1]$ is the probability of missingness.

For a given sample (x, y) , let x_{obs} denote the observed portion of x and θ_{obs} denote the corresponding sub-vector of θ . Define the missing portions x_{mis} to be the missing portion of x and θ_{mis} to be the corresponding sub-vector of θ .

W.L.O.G. we can assume that the coordinates are permuted and the missing values are in the first block. Under the framework of the Linear Regression with Missing Covariates, the EM algorithm imputes the conditional mean and conditional covariance using the current parameter estimate θ . Immediately, the conditional mean of X given (x_{obs}, y) is

$$\mu_\theta(x_{\text{obs}}, y) := \begin{bmatrix} \mathbb{E}(x_{\text{mis}}|x_{\text{obs}}, y, \theta) \\ x_{\text{obs}} \end{bmatrix} = \begin{bmatrix} U_\theta z_{\text{obs}} \\ x_{\text{obs}} \end{bmatrix} \quad (73)$$

where

$$U_\theta = \frac{1}{\|\theta_{\text{mis}}\|_2^2 + \sigma^2} \begin{bmatrix} -\theta_{\text{mis}} \theta_{\text{obs}}^T & \theta_{\text{mis}} \end{bmatrix} \quad \text{and} \quad z_{\text{obs}} := \begin{bmatrix} x_{\text{obs}} \\ y \end{bmatrix} \in \mathbb{R}^{|x_{\text{obs}}|+1} \quad (74)$$

Similarly, for the second moment matrix, we have

$$\Sigma_\theta(x_{\text{obs}}, y) := \mathbb{E}[XX^T|x_{\text{obs}}, y, \theta] = \begin{bmatrix} I & U_\theta z_{\text{obs}} x_{\text{obs}}^T \\ x_{\text{obs}} z_{\text{obs}}^T U_\theta^T & x_{\text{obs}} x_{\text{obs}}^T \end{bmatrix} \quad (75)$$

For a given parameter θ , the EM update is to maximize

$$Q_n(\theta'|\theta) := -\frac{1}{2n} \sum_{i=1}^n \langle \theta', \Sigma_\theta(x_{\text{obs},i}, y_i) \theta' \rangle + \frac{1}{n} \sum_{i=1}^n y_i \langle \mu_\theta(x_{\text{obs},i}, y_i), \theta' \rangle \quad (76)$$

The sample-based EM iterations are given as

$$\theta^{t+1} := \left[\sum_{i=1}^n \Sigma_{\theta^t}(x_{\text{obs},i}, y_i) \right]^{-1} \left[\sum_{i=1}^n y_i \mu_{\theta^t}(x_{\text{obs},i}, y_i) \right] \quad (77)$$

and the population EM iterations are given as

$$\theta^{t+1} := \{\mathbb{E}[\Sigma_{\theta^t}(X_{\text{obs}}, Y)]\}^{-1} \mathbb{E}[Y \mu_{\theta^t}(X_{\text{obs}}, Y)] \quad (78)$$

On the other hand, the sample-based first-order EM algorithm with step size α performs

$$\theta^{t+1} = \theta^t + \alpha \left\{ \frac{1}{n} \sum_{i=1}^n [y_i \mu_{\theta^t}(x_{\text{obs},i}, y_i) - \Sigma_{\theta^t}(x_{\text{obs},i}, y_i) \theta^t] \right\} \quad (79)$$

and the population-based first order EM algorithm with step size α performs

$$\theta^{t+1} = \theta^t + \alpha \mathbb{E}[Y \mu_{\theta^t}(X_{\text{obs}}, Y) - \Sigma_{\theta^t}(X_{\text{obs}}, Y) \theta^t] \quad (80)$$

B.3.1 Population Analysis

In this example, we also give some conditions on the signal-to-noise ratio and the radius of contractivity r (the radius of the region around θ^* within which the population EM algorithm is convergent to a global optimum). Define

$$\xi_1 := \frac{\|\theta^*\|_2}{\sigma} \quad \text{and} \quad \xi_2 := \frac{r}{\sigma} \quad (81)$$

The condition for ρ in our first corollary is given by

$$\rho < \frac{1}{1 + 2\xi(1 + \xi)} \quad \text{where } \xi := (\xi_1 + \xi_2)^2 \quad (82)$$

Corollary B.5 (Population contractivity for missing covariates). *Given any missing covariate regression model with missing probability ρ satisfying the bound (82) the first-order EM iterates with stepsize 1, satisfy the bound*

$$\|\theta^t - \theta^*\|_2 \leq \kappa^t \|\theta^0 - \theta^*\|_2 \quad \text{for } t = 1, 2, \dots \quad (83)$$

where $\kappa \equiv \kappa(\xi, \rho) := \left(\frac{\xi + \rho(1 + 2\xi(1 + \xi))}{1 + \xi} \right)$.

Proof. As before, we need to verify three conditions (μ -smooth, λ -strongly concave and that the GS condition) are satisfied for the function q which takes the form

$$q(\theta) = \frac{1}{2} \langle \theta, \mathbb{E}[\Sigma_{\theta^*}(X_{\text{obs}}, Y)] \theta \rangle - \langle \mathbb{E}[Y \mu_{\theta^*}(X_{\text{obs}}, Y)], \theta \rangle$$

where the vector $\mu_{\theta^*} \in \mathbb{R}^d$.

- Smoothness and strong concavity.

Note that

$$\nabla^2 q(\theta) = \mathbb{E}[\Sigma_{\theta^*}(X_{\text{obs}}, Y)]$$

By fixing a pattern of missingness and then averaging over (X_{obs}, Y) , it is easy to verify that

$$\mathbb{E}[\Sigma_{\theta^*}(X_{\text{obs}}, Y)] = \begin{bmatrix} I & U_{\theta^*} \begin{bmatrix} I \\ \theta_{\text{obs}}^{*T} \end{bmatrix} \\ \begin{bmatrix} I & \theta_{\text{obs}}^* \end{bmatrix} U_{\theta^*}^T & I \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}$$

Therefore, smoothness and strong concavity hold with $\mu = \lambda = 1$.

- GS Condition.

To verify this condition, we need to show that there is a scalar $\gamma \in [0, 1]$ such that

$$\|\mathbb{E}[V]\|_2 \leq \gamma \|\theta - \theta^*\|_2$$

where the vector

$$V = V(\theta, \theta^*) = \Sigma_{\theta^*}(X_{\text{obs}}, Y)\theta - Y\mu_{\theta^*}(X_{\text{obs}}, Y) - \Sigma_\theta(X_{\text{obs}}, Y)\theta + Y\mu_\theta(X_{\text{obs}}, Y) \quad (84)$$

We can compute the expectation over (X_{obs}, Y) assuming that the first block is missing as below

$$\mathbb{E}_{X_{\text{obs}}, Y}[V] = \begin{bmatrix} (\theta_{\text{mis}} - \theta_{\text{mis}}^*) + \pi_1 \theta_{\text{mis}} \\ \pi_2 (\theta_{\text{obs}} - \theta_{\text{obs}}^*) \end{bmatrix} \quad (85)$$

where

$$\begin{aligned}\pi_1 &:= \frac{\|\theta_{\text{mis}}^*\|_2^2 - \|\theta_{\text{mis}}\|_2^2 + \|\theta_{\text{obs}} - \theta_{\text{obs}}^*\|_2^2}{\|\theta_{\text{mis}}\|_2^2 + \sigma^2} \\ \pi_2 &:= \frac{\|\theta_{\text{mis}}\|_2^2}{\|\theta_{\text{mis}}\|_2^2 + \sigma^2}\end{aligned}$$

Claim B.11. π_1 and π_2 can be bounded independently of the missingness pattern by

$$\pi_1 \leq 2(\xi_1 + \xi_2) \frac{\|\theta - \theta^*\|_2}{\sigma} \quad \text{and} \quad \pi_2 \leq \delta := \frac{1}{1 + (1/(\xi_1 + \xi_2))^2} < 1 \quad (86)$$

To prove this claim, note that $\|\theta_{\text{mis}}\|_2 - \|\theta_{\text{mis}}^*\|_2 \leq \|\theta_{\text{mis}} - \theta_{\text{mis}}^*\|_2$ thus, by assumption, we have

$$\|\theta_{\text{mis}}\|_2 \leq \|\theta_{\text{mis}}^*\|_2 + \xi_2 \sigma \leq (\xi_1 + \xi_2) \sigma \quad (87)$$

and thus

$$\begin{aligned}\|\theta_{\text{mis}}^*\|_2^2 - \|\theta_{\text{mis}}\|_2^2 &= (\|\theta_{\text{mis}}\|_2 - \|\theta_{\text{mis}}^*\|_2)(\|\theta_{\text{mis}}\|_2 + \|\theta_{\text{mis}}^*\|_2) \\ &\leq (2\xi_1 + \xi_2) \sigma \|\theta_{\text{mis}} - \theta_{\text{mis}}^*\|_2\end{aligned}$$

Therefore, the bound for π_1 follows from $\|\theta_{\text{obs}} - \theta_{\text{obs}}^*\|_2^2 \leq \xi_2 \sigma \|\theta_{\text{obs}} - \theta_{\text{obs}}^*\|_2$. Besides,

$$\pi_2 = \frac{\|\theta_{\text{mis}}\|_2^2}{\|\theta_{\text{mis}}\|_2^2 + \sigma^2} = \frac{1}{1 + \sigma^2/\|\theta_{\text{mis}}\|_2^2} \stackrel{(i)}{=} \frac{1}{1 + (1/(\xi_1 + \xi_2))^2} < 1$$

where the first inequality follows from equation (87)

Now using the results in the claim above, we can then average over the missing pattern. Note that each coordinate is missing independently with probability ρ , we have

$$|\mathbb{E}[V]|_i \leq |\rho|\theta_i - \theta_i^*| + \rho\pi_1|\theta_i| + (1-\rho)\pi_2|\theta_i - \theta_i^*|$$

Let $\eta := (1-\rho)\delta + \rho < 1$, we have

$$\begin{aligned}\|\mathbb{E}[V]\|_2^2 &\leq \eta^2 \|\theta - \theta^*\|_2^2 + \rho^2 \pi_1^2 \|\theta\|_2^2 + 2\pi_1 \eta \rho \|\theta - \theta^*\| \\ &\leq \left\{ \eta^2 + \rho^2 \|\theta\|_2^2 \frac{4(\xi_1 + \xi_2)^2}{\sigma^2} + \frac{4\eta\rho \|\theta\|_2 (\xi_1 + \xi_2)}{\sigma} \right\} \|\theta - \theta^*\|_2^2\end{aligned}$$

Define

$$\gamma^2 := \eta^2 + \rho^2 \|\theta\|_2^2 \frac{4(\xi_1 + \xi_2)^2}{\sigma^2} + \frac{4\eta\rho \|\theta\|_2 (\xi_1 + \xi_2)}{\sigma}$$

By assumption, we have $\|\theta^*\|_2 \leq \xi_1 \sigma$ and $\|\theta - \theta^*\|_2 \leq \xi_2 \sigma$, and hence $\|\theta\|_2 \leq (\xi_1 + \xi_2) \sigma$. Thus, the coefficient γ^2 can be bounded as

$$\gamma^2 \leq \eta^2 + 4\rho^2 (\xi_1 + \xi_2)^4 + 4\eta\rho (\xi_1 + \xi_2)^2$$

Under the conditions of the corollary, we have $\gamma < 1$ which completes the proof. \square

B.3.2 Sample-based Analysis

Corollary B.6 (Sample-splitting first-order EM guarantees for missing covariates). *In addition to the conditions of Corollary B.5 suppose that the sample size is lower bounded as $n \geq c_1 d \log(1/\delta)$. Then there is a contraction coefficient $\kappa < 1$ such that, for any initial vector $\theta^0 \in \mathbb{B}_2(\xi_2\sigma; \theta^*)$, the sample-splitting first-order EM iterates with step size 1, based on n/T samples per iteration satisfy the bound*

$$\|\theta^t - \theta^*\|_2 \leq \kappa^t \|\theta^0 - \theta^*\|_2 + \frac{c_2 \sqrt{1 + \sigma^2}}{1 - \kappa} \sqrt{\frac{d}{n} T \log(T/\delta)} \quad (88)$$

with probability at least $1 - \delta$

Proof. Similar to Corollary B.4, our goal is to bound $\varepsilon_Q(n, \delta)$ defined as

$$\mathbb{P} \left[\left\| \nabla Q_n(\theta | \theta^t) \Big|_{\theta=\theta^t} - \nabla Q(\theta | \theta^t) \Big|_{\theta=\theta^t} \right\|_2 > \varepsilon_Q(n, \delta) \right] \leq 1 - \delta$$

For any fixed $\theta \in \mathbb{B}_2(r; \theta^*) = \{\theta \in \mathbb{R}^d \mid \|\theta - \theta^*\|_2 \leq \xi_2\sigma\}$, we need to upper bound the random variable,

$$Z = \left\| \frac{1}{n} \sum_{i=1}^n [y_i \mu_\theta(x_{\text{obs},i}, y_i) - \Sigma_\theta(x_{\text{obs},i}, y_i) \theta] - \mathbb{E}[Y \mu_\theta(X_{\text{obs}}, Y) - \Sigma_\theta(X_{\text{obs}}, Y) \theta] \right\|_2$$

with high probability. We define:

$$\begin{aligned} T_1 &:= \left\| \left[\mathbb{E}\Sigma_\theta(x_{\text{obs}}, y) \theta - \frac{1}{n} \sum_{i=1}^n \Sigma_\theta(x_{\text{obs},i} + y_i) \theta \right] \right\|_2 \\ T_2 &:= \left\| \left[\mathbb{E}(y \mu_\theta(x_{\text{obs}}, y)) - \frac{1}{n} \sum_{i=1}^n y_i \mu_\theta(x_{\text{obs},i}, y_i) \right] \right\|_2 \end{aligned}$$

Let $z_i \in \mathbb{R}^d$ be i -th canonical vector with ones in the positions of observed covariates and the notation \odot be the element-wise product. Now we bound T_1 and T_2 respectively.

- Controlling T_1 : Define

$$\begin{aligned} \bar{\Sigma} &= \mathbb{E}[\Sigma_\theta(x_{\text{obs}}, y)] \\ \hat{\Sigma} &= \frac{1}{n} \sum_{i=1}^n \Sigma_\theta(x_{\text{obs},i}, y_i) \end{aligned}$$

Then we have

$$T_1 \leq \|\bar{\Sigma} - \hat{\Sigma}\|_{\text{op}} \|\theta\|_2 \leq \|\bar{\Sigma} - \hat{\Sigma}\|_{\text{op}} (\xi_1 + \xi_2) \sigma$$

where the second step follows since any vector $\theta \in \mathbb{B}_2(r; \theta^*)$ has ℓ_2 -norm bounded as $\|\theta\|_2 \leq (\xi_1 + \xi_2) \sigma$.

Claim B.12. For any fixed vector $u \in S^d$, the random variable $\langle u, (\bar{\Sigma} - \hat{\Sigma})u \rangle$ is zero-mean and sub-exponential.

The key idea is to use the expression

$$\Sigma_\theta(x_{\text{obs}}, y) = I_{\text{mis}} + \mu_\theta \mu_\theta^T - ((1 - z) \odot \mu_\theta) ((1 - z) \odot \mu_\theta)^T$$

where I_{mis} is the identity matrix on the diagonal sub-block corresponding to the missing entries. Since the square of any sub-Gaussian random variable has sub-exponential tails. Thus, we only need to show that each of the random variables $\langle \mu_\theta, u \rangle$, and $\langle (1 - z) \odot \mu_\theta, u \rangle$ are sub-Gaussian. To show this, we need to verify that μ_θ is sub-Gaussian and here we state this argument as following Lemma

Lemma B.4. Under the conditions of Corollary B.5, the random vector $\mu_\theta(x_{\text{obs}}, y)$ is sub-Gaussian with a constant parameter.

For the detailed proof of this lemma and the claim above, you can refer to Page 115 in Statistical Guarantees for the EM Algorithm: From Population to Sample-based Analysis.

By above claim and referring to some standard arguments in random matrix theory, we have

$$\|\bar{\Sigma} - \hat{\Sigma}\|_{\text{op}} \leq c \sqrt{\frac{d}{n} \log(1/\delta)}$$

with probability at least $1 - \delta$ when $n > d$.

- Controlling T_2 : Note that

$$T_2 = \sup_{\|u\|_2=1} \left| \mathbb{E}[y \langle \mu_\theta(x_{\text{obs}}, y), u \rangle] - \frac{1}{n} \sum_{i=1}^n y_i \langle \mu_\theta(x_{\text{obs},i}, y_i), u \rangle \right|$$

Using the discretization argument with a $1/2$ -cover $\{u^1, \dots, u^M\}$ of the sphere with $M \leq 2^d$ elements gives us

$$T_2 \leq 2 \max_{j \in [M]} \left| \mathbb{E}[y \langle \mu_\theta(x_{\text{obs}}, y), u^j \rangle] - \frac{1}{n} \sum_{i=1}^n y_i \langle \mu_\theta(x_{\text{obs},i}, y_i), u^j \rangle \right|$$

Each term in this maximum is the product of two zero-mean variables. To bound T_2 , we have following observations

- (i) y is sub-Gaussian with parameter at most $\sqrt{\|\theta^*\|_2^2 + \sigma^2} \leq c\sigma$;
- (ii) $\langle \mu_\theta, u \rangle$ is sub-Gaussian with constant parameter by Lemma B.4.
- (iii) The product of any two sub-Gaussian variables is sub-exponential.

Therefore, by standard sub-exponential tail bounds, we have

$$\mathbb{P}[T_2 \geq t] \leq 2M \exp\left(-c \min\left\{\frac{nt}{\sqrt{1+\sigma^2}}, \frac{nt^2}{1+\sigma^2}\right\}\right)$$

Since $M \leq 2^d$ and $n > c_1 d$, we have

$$T_2 \leq c \sqrt{1+\sigma^2} \sqrt{\frac{d}{n} \log(1/\delta)}$$

with probability at least $1 - \delta$.

Combining our bounds on T_1 and T_2 , we have

$$\varepsilon_Q(n, \delta) \leq c \sqrt{1+\sigma^2} \times \sqrt{\frac{d}{n} \log(1/\delta)}$$

with probability at least $1 - \delta$ and the Corollary B.6 follows from Theorem 2. \square

Remark B.4. If we set $T = c \log n$ for a sufficiently large constant c , then the bound (88) implies that

$$\|\theta^T - \theta^*\|_2 \leq c' \sqrt{1+\sigma^2} \sqrt{\frac{d}{n} \log^2(n/\delta)}$$

with probability at least $1 - \delta$.

C Additional Experiment

Besides looking at the optimization error for different SNR for Mixture of Gaussians, using EM, we did the following additional experiments: First-Order-EM, for Mixture of Gaussians (Fig 2); EM, for Mixture of Regressions (Fig 3); First-Order-EM, for Mixture of Regressions (Fig 4).

All of them confirmed that the optimization error decreased faster with larger values of signal-to-noise ratio.

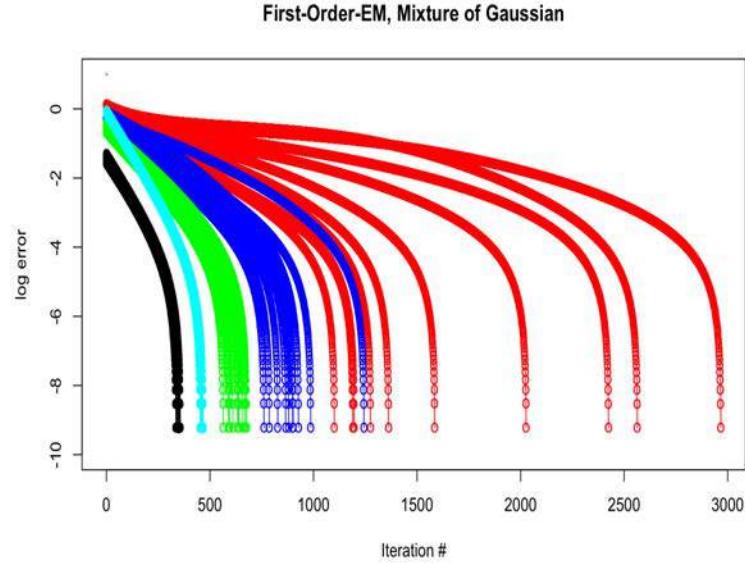


Figure 2: First-Order-EM: Mixture of Gaussians for Different SNR

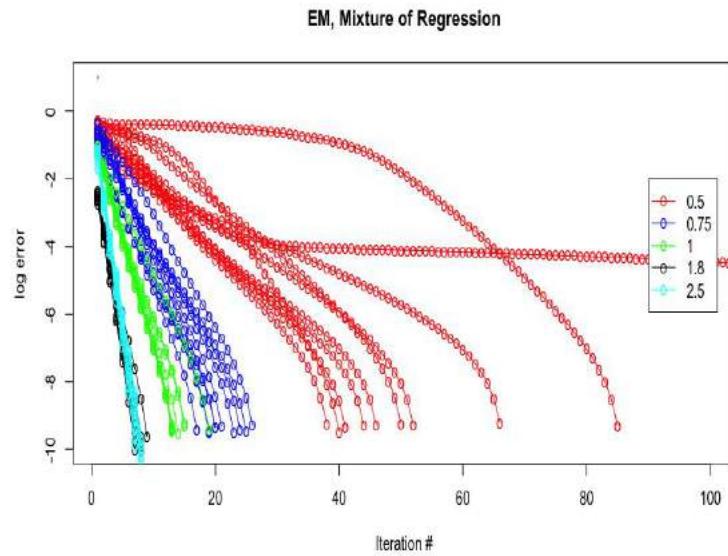


Figure 3: EM: Mixture of Regressions for Different SNR

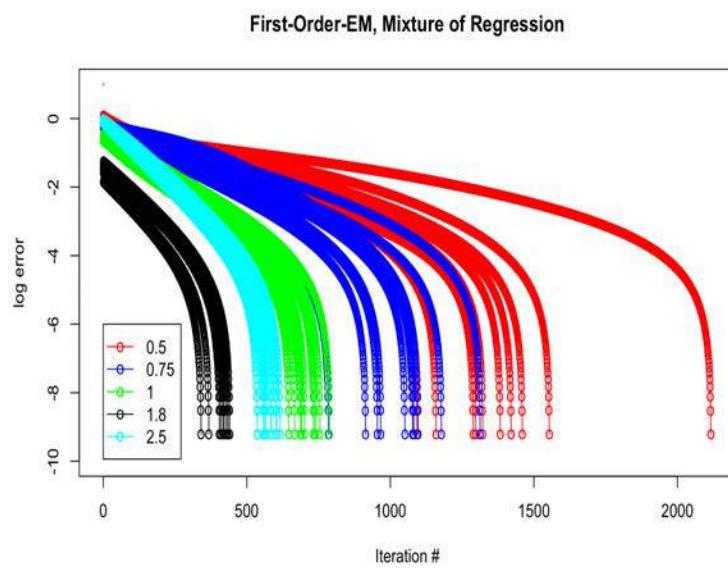


Figure 4: First-Order-EM: Mixture of Regressions for Different SNR

D R Code For Experiments

D.1 Mixture of Gaussians

```
## EM-Algorithm for GMM
#X: the data matrix, n X d. n is the number of observations, d is the dimension of
#features.
#eta: the tolerance of difference of the theta values.
EMGMM<-function(data,theta0,sigma0,alpha=0.01,eta=0.00001){
  X=data$X
  sigma20=sigma0^2
  n = nrow(X); d = ncol(X)
  theta.t = list()
  repeat({
    first.term=0
    second.term=0
    for(i in 1:n){
      xi=X[i,]
      w.n=exp(-norm(theta0-xi,"2")^2/(2*sigma20))
      w.p=exp(-norm(theta0+xi,"2")^2/(2*sigma20))
      w=w.n*(w.n+w.p)^(-1)
      first.term=first.term+w*xi
      second.term=second.term+xi
    }
    theta=2/n*first.term-1/n*second.term
    err = norm(theta - theta0, "2")
    theta.t[[length(theta.t)+1]] = theta
    theta0=theta
    #print(err)
    if(err<eta)
      break
  })
  return(list(theta.hat=theta,theta.t=theta.t))

## First-Order EM-Algorithm for GMM
FEMGMM<-function(data,theta0,sigma0,alpha=0.01,eta=0.0001){
  X=data$X
  sigma20=sigma0^2
  n = nrow(X); d = ncol(X)
  theta.t = list()
  repeat({
    gradient=0
    for(i in 1:n){
      xi=X[i,]
      w.n=exp(-norm(theta0-xi,"2")^2/(2*sigma20))
      w.p=exp(-norm(theta0+xi,"2")^2/(2*sigma20))
      w=w.n*(w.n+w.p)^(-1)
      gradient=gradient+(2*w-1)*xi
    }
    update=1/n*gradient-theta0
    theta=theta0+alpha*update
    err = norm(theta - theta0, "2")
    theta.t[[length(theta.t)+1]] = theta
    theta0=theta
    #print(err)
```

```

    if(err<eta)
      break
  })
  return(list(theta.hat=theta,theta.t=theta.t))
}

## GMM Data Generation
GMMdata<-function(c,sigma){
  # c decide the norm of the theta.opt
  n = 1000;k=2;d=10;sigma2 = sigma^2;
  gauss = rnorm(d)
  length = norm(gauss,'2')
  c=c
  ## construct the theta star that length is c
  theta.opt=c*gauss/length

  theta = matrix(1:(k*d),nrow=k,byrow = T)
  theta[1,] = theta.opt
  theta[2,] = -theta.opt
  pi = c(1,1)/k
  Z = sample(1:k,n,prob=pi,replace = T)
  X = matrix(0,nrow=n,ncol=d)
  for(i in 1:n){
    X[i,] = rnorm(d,mean=theta[Z[i],],sd=sqrt(sigma2))
  }
  return(list(X=X,theta=theta.opt))
}

```

D.2 Mixture of Regressions

```

## EM for Mixture of Regressions
EMMR<-function(data,theta0,sigma0,alpha=0.01,eta=0.0001){
  X=data$X
  y=data$y
  sigma20=sigma0^2
  n = nrow(X); d = ncol(X)
  theta.t=list()
  repeat({
    first.term=0
    second.term=0
    for(i in 1:n){
      xi=X[i,]
      yi=y[i,]
      w.n=exp(-(yi-xi%*%theta0)^2/(2*sigma20))
      w.p=exp(-(yi+xi%*%theta0)^2/(2*sigma20))
      w=w.n*(w.n+w.p)^(-1)
      first.term=first.term+xi%*%t(xi)
      second.term=second.term+(2*w-1)*yi*xi
    }
    theta=solve(first.term)%*%second.term
    err = norm(theta - theta0,"2")
    theta.t[[length(theta.t)+1]] = theta
    theta0=theta
  })
}
```

```

#print(err)
if(err<eta)
  break
})
return(list(theta.hat=theta,theta.t=theta.t))
}

## first-order EM for Mixture of Regressions
FEMMR<-function(data,theta0,sigma0,alpha=0.01,eta=0.0001){
  X=data$X
  y=data$y
  sigma20=sigma0^2
  n = nrow(X); d = ncol(X)
  theta.t=list()
  repeat({
    gradient=0
    for(i in 1:n){
      xi=X[i,]
      yi=y[i,]
      w.n=exp(-(yi-xi%*%theta0)^2/(2*sigma20))
      w.p=exp(-(yi+xi%*%theta0)^2/(2*sigma20))
      w=w.n/(w.n+w.p)
      gradient=gradient+(2*w-1)*yi*xi-xi%*%t(xi)%*%theta0
    }
    update=1/n*gradient
    theta=theta0+alpha*update
    err = norm(theta - theta0, "2")
    theta.t[[length(theta.t)+1]] = theta
    theta0=theta
    #print(err)
    if(err<eta)
      break
  })
  return(list(theta.hat=theta,theta.t=theta.t))
}

## Mixture of regression data
MRdata<-function(c,sigma){
  sigma2 = sigma^2
  n = 1000;k=2;d=10;sigma2 = 1;
  gauss = rnorm(d)
  length = norm(gauss,'2')
  c=c
  theta.opt=c*gauss/length
  theta = matrix(1:(k*d),nrow=k,byrow = T)
  theta[1,] = theta.opt
  theta[2,] = -theta.opt
  pi = c(1,1)/k
  Z = sample(1:k,n,prob=pi,replace = T)
  X = matrix(0,nrow=n,ncol=d)
  y = matrix(0,nrow=n, ncol=1)
  for(i in 1:n){
    xi = rnorm(d,mean=0,sd=1)
    vi = rnorm(1, mean=0, sd=1)

```

```

X[i,] = xi
y[i,] = xi%*%theta[Z[i],]+vi
}
return(list(X=X,y=y,theta=theta.opt))
}

```

D.3 Linear Regression with Missing Covariates

```

## EM for LRMC
EMLRMC<-function(data,theta0,sigma0,alpha=0.01,eta=0.0001){
  X=data$X
  y=data$y
  sigma20=sigma0^2
  n = nrow(X); d = ncol(X)
  theta.t = list()
  repeat({
    first.term=0
    second.term=0
    for(i in 1:n){
      xi=X[i,]
      yi=y[i,]
      NAindex <-which(is.na(xi))
      numNA = length(NAindex)

      if(numNA>0)
      {
        NonNAindex <- which(!is.na(xi))
        xi.obs=xi[NonNAindex]
        xi.mis=xi[NAindex]
        theta0.obs=theta0[NonNAindex]
        theta0.mis=theta0[NAindex]
        z.obs=c(xi.obs,yi)
        U.theta=cbind(-theta0.mis%*%t(theta0.obs),theta0.mis)*1
          /(norm(theta0.mis,type="2")^2+sigma20)
        u=c(U.theta%*%z.obs,xi.obs)
        p1=matrix(1, numNA, numNA)
        p2=U.theta%*%z.obs%*%t(xi.obs)
        p3=xi.obs%*%t(z.obs)%*%t(U.theta)
        p4=xi.obs%*%t(xi.obs)
        #conditional second moment matrix
        csmm=rbind(cbind(p1,p2),cbind(p3,p4))
        second.term=second.term+yi*u
        first.term=first.term+csmm
      }
      else
      {
        second.term=second.term+yi*xi #d*1 vector
        first.term=first.term+xi%*%t(xi) #d*d matrix
      }
    }
    theta=solve(first.term)%*%second.term
    err = norm(theta - theta0,"2")
  })
}
```

```

theta.t[[length(theta.t)+1]] = theta
theta0=theta
#print(err)
if(err<eta)
  break
})
return(list(theta.hat=theta,theta.t=theta.t))
}

## First-Order-EM for LRMC

FEMLRMC<-function(data,theta0,sigma0,alpha=0.01,eta=0.0001){
X=data$X
y=data$y
sigma20=sigma0^2
n = nrow(X); d = ncol(X)
theta.t = list()
repeat({
gradient=0
for(i in 1:n){
  xi=X[i,]
  yi=y[i,]
  NAindex <-which(is.na(xi))
  numNA = length(NAindex)
  if(numNA>0)
  {
    NonNAindex <- which(!is.na(xi))
    xi.obs=xi[NonNAindex]
    xi.mis=xi[NAindex]
    theta0.obs=theta0[NonNAindex]
    theta0.mis=theta0[NAindex]
    z.obs=c(xi.obs,yi)
    U.theta=cbind(-theta0.mis%*%t(theta0.obs),theta0.mis)*1
      /(norm(theta0.mis,type="2")^2+sigma20)
    u=c(U.theta%*%z.obs,xi.obs)
    p1=matrix(1, numNA, numNA)
    p2=U.theta%*%z.obs%*%t(xi.obs)
    p3=xi.obs%*%t(z.obs)%*%t(U.theta)
    p4=xi.obs%*%t(xi.obs)
    #conditional second moment matrix
    csmm=rbind(cbind(p1,p2),cbind(p3,p4))
    gradient=gradient+yi*u-csmm%*%theta0
  }
  else
  {
    gradient=gradient+yi*xi-xi%*%t(xi)%*%theta0
  }
}
update=1/n*gradient
theta=theta0+alpha*update
err = norm(theta - theta0,"2")
theta.t[[length(theta.t)+1]] = theta
}

```

```

    theta0=theta
    #print(err)
    if(err<eta)
      break
  })
  return(list(theta.hat=theta,theta.t=theta.t))
}

## Linear Regression with Missing Covariates data generation

LRMCdata<-function(c,sigma){
  n = 1000;k=2;d=10;sigma2 = sigma^2;p = 0.2
  y = matrix(0,nrow=n, ncol=1)
  gauss = rnorm(d)
  length = norm(gauss,'2')
  c=c
  theta.opt=c*gauss/length
  #print(norm(theta.opt,'2'))
  theta = matrix(1:(k*d),nrow=k,byrow = T)
  theta[1,] = theta.opt
  theta[2,] = -theta.opt
  pi = c(1,1)/k
  Z = sample(1:k,n,prob=pi,replace = T)
  X = matrix(0,nrow=n,ncol=d)
  y = matrix(0,nrow=n, ncol=1)
  for(i in 1:n){
    xi = rnorm(d,mean=0,sd=1)
    mask = sample(c(NA,1), prob=c(p, 1-p), replace=TRUE, size=d)
    X[i,] = xi*mask
    vi = rnorm(1, mean=0,sd=1)
    y[i,] = sum(X[i,]*theta[Z[i],], na.rm = TRUE) +vi
  }
  return(list(X=X,y=y,theta=theta.opt))
}

```

D.4 Experiments

```

## function to run 10 times for a given theta.opt and sigma and given algorithm
r1 = function(c,sigma,datafunc,algofunc){
  data = datafunc(c,sigma)
  theta.opt = data$theta
  theta0=c()
  ## based on Fig3, the theta.opt_i - sqrt(c/10)<=theta0_i<=theta.opt_i + sqrt(c/10)
  dif = runif(1,c*sqrt(1/10)-0.5,c*sqrt(1/10)-0.5)
  for(i in 1:length(theta.opt)){
    p = sample(c(1,-1),1)
    value = theta.opt[i] + p*dif
    theta0 = append(theta0, value)
  }
  norm(theta.opt,'2')
  norm(theta0,'2')
  norm((theta.opt-theta0),'2')
  # call the function
}

```

```

mod=algorfunc(data,theta0,sigma)
theta.t = mod$theta.t
theta.hat = mod$theta.hat
end = length(theta.t)

print(end)
err.opt = c()
err.stat = c()
for(i in 1:end){
  theta = theta.t[[i]]
  e1=log(norm(theta.hat-theta,'2'))
  e2=log(norm(theta.opt-theta,'2'))
  #print(e1)
  #print(e2)
  err.opt=c(err.opt, e1)
  err.stat=c(err.stat, e2)
}
return(list(err.opt=err.opt,err.stat=err.stat))
}

## Fig5(a)
rs = lapply(rep(2,10), function(c){
  sigma = 1
  out=r1(c,sigma,GMMdata,EMGMM)
})

end = 0
for(i in 1:10){
  err.opt=rs[[i]]$err.opt
  end_ = length(err.opt)
  end = max(end,end_)
}

x = seq(1:end-1)
plot(x,xlab="Iteration #", ylab="log error",ylim=c(-8,1),
     main='EM, Mixture of Gaussian',cex=1,lwd=1)

for(i in 1:10){
  err.opt=rs[[i]]$err.opt
  err.stat=rs[[i]]$err.stat
  end = length(err.opt)-1
  x = seq(1:end)
  points(x, err.opt[1:end], col="blue", pch="o",cex=1,lwd=1)
  lines(x, err.opt[1:end], col="blue", lty=1, cex=1,lwd=1)
  points(x, err.stat[1:end], col="red", pch="*",cex=1,lwd=1)
  lines(x, err.stat[1:end], col="red", lty=2,cex=1,lwd=1)
}
legend(1,-6,legend=c("Opt. error","Stat. error"), col=c("blue","red"),
       pch=c("o","*"),lty=c(1,2), ncol=1)

## Fig 5(b)
rs = lapply(rep(2,10), function(c){
  sigma = 1
  out=r1(c,sigma,GMMdata,FEMGMM)
})

```

```

})

end = 0
for(i in 1:10){
  err.opt=rs[[i]]$err.opt
  end_ = length(err.opt)
  end = max(end,end_)
}

end=500
x = seq(1:end-1)
plot(x,xlab="Iteration #", ylab="log error",ylim=c(-6,1),
     main='First-order EM, Mixture of Gaussian',cex=0.3,lwd=0.3)

for(i in 1:10){
  err.opt=rs[[i]]$err.opt
  err.stat=rs[[i]]$err.stat
  end = length(err.opt)-1
  end=500
  x = seq(1:end)
  points(x, err.opt[1:end], col="blue", pch="o",cex=0.3,lwd=0.3)
  lines(x, err.opt[1:end], col="blue", lty=1, cex=0.3,lwd=0.3)
  points(x, err.stat[1:end], col="red", pch="*",cex=0.3,lwd=0.3)
  lines(x, err.stat[1:end], col="red", lty=2,cex=0.3,lwd=0.3)
}
legend(1,-4,legend=c("Opt. error","Stat. error"), col=c("blue","red"),
       pch=c("o","*"),lty=c(1,2), ncol=1)

## Fig6
end = 0
rss = list()
snrs = c(0.5,0.75,1,1.8,2.5)
repnum=10
for(i in snrs){
  rs = lapply(rep(i,repnum), function(i){
    sigma = 1
    out=r1(i,sigma,GMMdata,EMGMM)
  })
  for(j in 1:repnum){
    err.opt=rs[[j]]$err.opt
    end_ = length(err.opt)
    end = max(end,end_)
  }
  rss[[length(rss)+1]]=rs
}

x = seq(1:end+1)
plot(x,xlab="Iteration #", ylab="log error",ylim=c(-10,1),
     main='EM, Mixture of Gaussian',cex=0.3,lwd=0.3)

colors=c('red','blue','green','black','cyan')
for(j in 1:length(snrs)){
  rs=rss[[j]]

```

```

color=colors[j]
for(i in 1:repnum){
  err.opt=rs[[i]]$err.opt
  end = length(err.opt)-1
  x = seq(1:end)
  points(x, err.opt[1:end], col=color, pch="o",cex=1,lwd=1)
  lines(x, err.opt[1:end], col=color, lty=1, cex=1,lwd=1)
}
legend(70,-2,legend=snrs, col=colors,
       pch=rep("o",5),lty=rep(1,5), ncol=1)

## Fig7(a)
## based on the theorem, the theta.opt_i - sqrt(c/10)<=theta0_i<=theta.opt_i + sqrt(c/10)
rs = lapply(rep(2,10), function(c){
  sigma = 1
  out=r1(c,sigma,MRdata,EMMR)
})

end = 0
for(i in 1:10){
  err.opt=rs[[i]]$err.opt
  end_ = length(err.opt)
  end = max(end,end_)
}

x = seq(1:end-1)
plot(x,xlab="Iteration #", ylab="log error",ylim=c(-8,1),
      main='EM, Mixture of Regression',cex=1,lwd=1)

for(i in 1:10){
  err.opt=rs[[i]]$err.opt
  err.stat=rs[[i]]$err.stat
  end = length(err.opt)-1
  x = seq(1:end)
  points(x, err.opt[1:end], col="blue", pch="o",cex=1,lwd=1)
  lines(x, err.opt[1:end], col="blue", lty=1, cex=1,lwd=1)
  points(x, err.stat[1:end], col="red", pch="*",cex=1,lwd=1)
  lines(x, err.stat[1:end], col="red", lty=2,cex=1,lwd=1)
}
legend(1,-6,legend=c("Opt. error","Stat. error"), col=c("blue","red"),
       pch=c("o","*"),lty=c(1,2), ncol=1)

## Fig 7(b)
rs = lapply(rep(2,10), function(c){
  sigma = 1
  out=r1(c,sigma,MRdata,FEMMR)
})

end = 0
for(i in 1:10){
  err.opt=rs[[i]]$err.opt

```

```

end_ = length(err.opt)
end = max(end,end_)
}

end=400
x = seq(1:end-1)
plot(x,xlab="Iteration #", ylab="log error",ylim=c(-8,1),
     main='First-order EM, Mixture of Regression',cex=0.3,lwd=0.3)

for(i in 1:10){
  err.opt=rs[[i]]$err.opt
  err.stat=rs[[i]]$err.stat
  end = length(err.opt)-1
  end=400
  x = seq(1:end)
  points(x, err.opt[1:end], col="blue", pch="o",cex=0.3,lwd=0.3)
  lines(x, err.opt[1:end], col="blue", lty=1, cex=0.3,lwd=0.3)
  points(x, err.stat[1:end], col="red", pch="*",cex=0.3,lwd=0.3)
  lines(x, err.stat[1:end], col="red", lty=2,cex=0.3,lwd=0.3)
}
legend(1,-6,legend=c("Opt. error","Stat. error"), col=c("blue","red"),
       pch=c("o","*"),lty=c(1,2), ncol=1)

## fig6 suplement for regression
end = 0
rss = list()
snrs = c(0.5,0.75,1,1.8,2.5)
repnum=10
for(i in snrs){
  rs = lapply(rep(i,repnum), function(i){
    sigma = 1
    out=r1(i,sigma,MRdata,EMMR)
  })
  for(j in 1:repnum){
    err.opt=rs[[j]]$err.opt
    end_ = length(err.opt)
    end = max(end,end_)
  }
  rss[[length(rss)+1]]=rs
}

end=100
x = seq(1:end+1)
plot(x,xlab="Iteration #", ylab="log error",ylim=c(-10,1),
      main='EM, Mixture of Regression',cex=0.3,lwd=0.3)

colors=c('red','blue','green','black','cyan')
for(j in 1:length(snrs)){
  rs=rss[[j]]
  color=colors[j]
  for(i in 1:repnum){
    err.opt=rs[[i]]$err.opt

```

```

    end=100
    end = length(err.opt)-1
    x = seq(1:end)
    points(x, err.opt[1:end], col=color, pch="o", cex=1, lwd=1)
    lines(x, err.opt[1:end], col=color, lty=1, cex=1, lwd=1)
  }
}
legend(90,-2,legend=snrs, col=colors,
       pch=rep("o",5),lty=rep(1,5), ncol=1)

## Fig8(a)
rs = lapply(rep(2,10), function(c){
  sigma = 1
  out=r1(c,sigma,LRMCdata,EMLRMC)
})

end = 0
for(i in 1:10){
  err.opt=rs[[i]]$err.opt
  end_ = length(err.opt)
  end = max(end,end_)
}

end=10
x = seq(1:end-1)
plot(x,xlab="Iteration #", ylab="log error",ylim=c(-7,1),
      main='EM, Missing Data Regression')

for(i in 1:10){
  err.opt=rs[[i]]$err.opt
  err.stat=rs[[i]]$err.stat
  end = length(err.opt)-1
  end=10
  x = seq(1:end)
  points(x, err.opt[1:end], col="blue", pch="o")
  lines(x, err.opt[1:end], col="blue", lty=1)
  points(x, err.stat[1:end], col="red", pch="*")
  lines(x, err.stat[1:end], col="red", lty=2)
}
legend(1,-5,legend=c("Opt. error","Stat. error"), col=c("blue","red"),
       pch=c("o","*"),lty=c(1,2), ncol=1)

## Fig8(b)
rs = lapply(rep(2,10), function(c){
  sigma = 1
  out=r1(c,sigma,LRMCdata,FEMLRMC)
})

end = 0
for(i in 1:10){
  err.opt=rs[[i]]$err.opt
  end_ = length(err.opt)
  end = max(end,end_)
}

```

```
}

end=500
x = seq(1:end+1)
plot(x,xlab="Iteration #", ylab="log error",ylim=c(-7.5,1),
     main='First-order EM, Missing Data Regression',cex=0.3,lwd=0.3)

for(i in 1:10){
  err.opt=rs[[i]]$err.opt
  err.stat=rs[[i]]$err.stat
  end = length(err.opt)-1
  end =500
  x = seq(1:end)
  points(x, err.opt[1:end], col="blue", pch="o",cex=0.3,lwd=0.3)
  lines(x, err.opt[1:end], col="blue", lty=1, cex=0.3,lwd=0.3)
  points(x, err.stat[1:end], col="red", pch="*",cex=0.3,lwd=0.3)
  lines(x, err.stat[1:end], col="red", lty=2,cex=0.3,lwd=0.3)
}
legend(1,-5,legend=c("Opt. error","Stat. error"), col=c("blue","red"),
       pch=c("o","*"),lty=c(1,2), ncol=1)
```

Review of Large-dimensional Central Limit Theorem with Fourth-moment Error Bounds on Convex Sets and Balls

Zhi Zhang*

1 Introduction

Let $\{\xi_i\}_{i=1}^n$ be a sequence of independent mean-zero random vectors in \mathbb{R}^d . Let $W = \sum_{i=1}^n \xi_i$ and $\Sigma = \text{Var}(W)$. It is well known that under finite third-moment conditions and for fixed dimension d , the distribution of W can be approximated by a Gaussian distribution with error rate $O(1/\sqrt{n})$. In this paper, we consider the approximation of probabilities of convex sets and Euclidean balls. For convex sets, Bentkus (2005) proved for the above W that if Σ is invertible and $Z \sim N(0, \Sigma)$, then

$$\sup_{A \in \mathcal{A}} |\mathbb{P}(W \in A) - \mathbb{P}(Z \in A)| \leq Cd^{1/4} \sum_{i=1}^n \mathbb{E} \left| \Sigma^{-1/2} \xi_i \right|^3 \quad (1)$$

where \mathcal{A} is the collection of all measurable convex sets in \mathbb{R}^d , C is an absolute constant and $|\cdot|$ denotes the Euclidean norm when applied to a vector.

The first main result is that up to a logarithmic factor,

$$\sup_{A \in \mathcal{A}} |\mathbb{P}(W \in A) - \mathbb{P}(Z \in A)| \leq \log Cd^{1/4} \left(\sum_{i=1}^n \mathbb{E} \left| \Sigma^{-1/2} \xi_i \right|^4 \right)^{1/2} \quad (2)$$

The bound (2) is optimal up to the $d^{1/4}$ and the logarithmic factors.

We then consider the Gaussian approximation on the class \mathcal{B} of all Euclidean balls, which is arguably most relevant for statistical applications, e.g., chi-square tests. We show that the factor $d^{1/4}$ in (2) can be removed if we replace \mathcal{A} with \mathcal{B} . Furthermore, we obtain an error bound that

*Department of Statistics, UC Davis, wwzzhan@ucdavis.edu. This is the project report for STA250 Special Topics in Statistics, Stein Methods

typically vanishes as long as $d = o(n)$. Incidentally, the requirement $d = o(n)$ is necessary for the validity of the Gaussian approximation on balls.

For a matrix M , we use $\|M\|_{H.S.}$ to denote its Hilbert-Schmidt norm. We use C to denote positive absolute constants which may differ in different expressions. For a vector $x \in \mathbb{R}^d$, we use $x_j, 1 \leq j \leq d$ to denote its components. For a sequence of vectors $x_i \in \mathbb{R}^d, 1 \leq i \leq n$, we use x_{ij} to denote the j th component of x_i for $1 \leq j \leq d$. Similarly, we write X_j and X_{ij} for the components of random vectors X and X_i , respectively.

1.1 Notations and Preliminary

Notation 1.1. For two vectors $x, y \in \mathbb{R}^d$, $x \cdot y$ denotes their inner product. For two $d \times d$ matrices M and N , we write $\langle M, N \rangle_{H.S.}$ for their Hilbert-Schmidt inner product.

$$\langle M, N \rangle_{H.S.} = \text{tr}(M^T N) = \sum_i \langle M e_i, N e_i \rangle$$

where $\{e_i : i \in I\}$ an orthonormal basis.

Notation 1.2. For real-valued functions on \mathbb{R}^d we will write $\partial_i f(x)$ for $\partial f(x)/\partial x_i, \partial_{ij} f(x)$ for $\partial^2 f(x)/(\partial x_i \partial x_j)$ and so forth.

Notation 1.3. We write ∇f and $\text{Hess } f$ for the gradient and Hessian matrix of f , respectively. In addition, we denote by $\nabla^r f(x)$ the r -th derivative of f at x regarded as an r -linear form: The value of $\nabla^r f(x)$ evaluated at $u_1, \dots, u_r \in \mathbb{R}^d$ is given by

$$\langle \nabla^r f(x), u_1 \otimes \cdots \otimes u_r \rangle = \sum_{j_1, \dots, j_r=1}^d \partial_{j_1, \dots, j_r} f(x) u_{1,j_1} \cdots u_{r,j_r}$$

When $u_1 = \cdots = u_r =: u$, we write $u_1 \otimes \cdots \otimes u_r = u^{\otimes r}$ for short.

Notation 1.4. For any r -linear form T , its injective norm is defined by

$$|T|_\vee := \sup_{|u_1|_\vee \wedge \dots \wedge |u_r|_\vee \leq 1} |\langle T, u_1 \otimes \cdots \otimes u_r \rangle|$$

For an $(r-1)$ -times differentiable function $h : \mathbb{R}^d \rightarrow \mathbb{R}$, we write

$$M_r(h) := \sup_{x \neq y} \frac{|\nabla^{r-1} h(x) - \nabla^{r-1} h(y)|_\vee}{|x - y|}$$

Remark 1.1. Note that $M_r(h) = \sup_{x \in \mathbb{R}^d} |\nabla^r h(x)|_\vee$ if h is r -times differentiable.

Finally, we refer to the following bound for derivatives of the d -dimensional standard normal density ϕ , which will be used several times in the following (cf. the inequality after Eq. (4.9) of Raič (2019 b)):

$$\int_{\mathbb{R}^d} |\langle \nabla^s \phi(z), u^{\otimes s} \rangle| dz \leq C_s |u|^s \quad \text{for any fixed integer } s \quad (3)$$

where C_s is a constant depending only on s .

2 Main Theorem

2.1 Approximation on Convex Sets

Theorem 2.1. *Let $\xi = \{\xi_i\}_{i=1}^n$ be a sequence of centered independent random vectors in \mathbb{R}^d with finite fourth moments and set $W = \sum_{i=1}^n \xi_i$. Assume $\text{Var}(W) = \Sigma$ and Σ is invertible. Let $Z \sim N(0, \Sigma)$ be a centered Gaussian vector in \mathbb{R}^d with covariance matrix Σ . Then,*

$$\sup_{A \in \mathcal{A}} |\mathbb{P}(W \in A) - \mathbb{P}(Z \in A)| \leq Cd^{1/4} \left(\sum_{i=1}^n \mathbb{E} |\Sigma^{-1/2} \xi_i|^4 \right)^{1/2} \left(\left| \log \left(\sum_{i=1}^n \mathbb{E} |\Sigma^{-1/2} \xi_i|^4 \right) \right| \vee 1 \right) \quad (4)$$

where \mathcal{A} is the collection of all measurable convex sets in \mathbb{R}^d .

Remark 2.1. The bound (4) is optimal up to the $d^{1/4}$ and the logarithmic factors.

Proposition 2.1. *There is an absolute constant $C_0 > 0$ such that, for sufficiently large n , we can construct centered i.i.d. random vectors ξ_1, \dots, ξ_n in \mathbb{R}^d with finite fourth moments (which may depend on n) satisfying $\text{Var}(W) = I_d$ and*

$$\sup_{A \in \mathcal{A}} |\mathbb{P}(W \in A) - \mathbb{P}(Z \in A)| \geq C_0 \left(\sum_{i=1}^n \mathbb{E} |\xi_i|^4 \right)^{1/2}$$

as long as $d \leq \sqrt{n}/\log n$.

Remark 2.2. Consider the situation where $\xi_i = X_i/\sqrt{n}$ and $\{X_1, X_2, \dots\}$ is a sequence of i.i.d. mean-zero random vectors in \mathbb{R}^d with $\text{Var}(X_i) = I_d$. In this setting, $\Sigma = I_d$, and for the d -vector X_i , we have

$$\mathbb{E} |X_i|^3 \propto d^{3/2} \quad \mathbb{E} |X_i|^4 \propto d^2$$

In this case, the right-hand side of (4) is of the order $O\left(\frac{d^{5/2}}{n}\right)^{1/2}$ up to a logarithmic factor. In contrast, the right-hand side of (1) is of the order $O\left(\frac{d^{7/2}}{n}\right)^{1/2}$. Therefore, subject to the requirement of the existence of the fourth moment, (4) is preferable to (1) in the large-dimensional setting where $d \rightarrow \infty$.

2.2 Approximation on Euclidean Balls

In this section, we show that the factor $d^{1/4}$ appearing on the right-hand side of (4) can be removed if we restrict the approximation to the class of balls. To facilitate the application to the bootstrap, here we do not assume W and Z have the same covariance matrix.

Theorem 2.2. *Let $\xi = \{\xi_i\}_{i=1}^n$ be a sequence of centered independent random vectors in \mathbb{R}^d with finite fourth moments and set $W = \sum_{i=1}^n \xi_i$. Let $Z \sim N(0, \Sigma)$ be a centered Gaussian vector in \mathbb{R}^d with covariance matrix Σ . Assume Σ is invertible. Then*

$$\sup_{A \in \mathcal{B}} |\mathbb{P}(W \in A) - \mathbb{P}(Z \in A)| \leq C\Psi(\delta(W, \Sigma)) \quad (5)$$

where $\Psi(x) = x(|\log x| \vee 1)$, \mathcal{B} is the set of all Euclidean balls in \mathbb{R}^d and

$$\delta(W, \Sigma) := \left\| I_d - \text{Var} \left(\Sigma^{-1/2} W \right) \right\|_{H.S.} + \left(\sum_{i=1}^n \mathbb{E} \left| \Sigma^{-1/2} \xi_i \right|^4 \right)^{1/2}$$

Remark 2.3. We can see that if $\text{Var}(W) = \Sigma$, then the typical order of the right-hand side of (5) is $O\left(\frac{d^2}{n}\right)^{1/2}$ up to a logarithmic factor. It has near-optimal dependence on n and converges to 0 if $d = o(\sqrt{n})$.

In the next result, we sacrifice the rate of n to obtain the optimal growth rate of $d = o(n)$ in terms of the dimension.

Theorem 2.3. *Let ξ, W and Z be as in Theorem 2.2. Assume $\text{tr}(\Sigma^2) > 0$. Then*

$$\sup_{A \in \mathcal{B}} |\mathbb{P}(W \in A) - \mathbb{P}(Z \in A)| \leq \frac{C}{\text{tr}(\Sigma^2)^{1/4}} \sqrt{\tilde{\delta}(W, \Sigma)} \quad (6)$$

where

$$\tilde{\delta}(W, \Sigma) := \|\Sigma - \text{Var}(W)\|_{H.S.} + \sum_{j=1}^d |\Sigma_{jj} - \text{Var}(W_j)| + \sqrt{\sum_{i=1}^n \mathbb{E} |\xi_i|^4} + \sum_{j=1}^d \sqrt{\sum_{i=1}^n \mathbb{E} [\xi_{ij}^4]}$$

Remark 2.4. Since $\mathbb{E} |\xi_i|^4 \leq d \sum_{j=1}^d \mathbb{E} \xi_{ij}^4$, if $\text{Var}(W) = \Sigma = I_d$, the right-hand side of (6) is bounded by

$$C \max_{1 \leq j \leq d} \left(d \sum_{i=1}^n \mathbb{E} \xi_{ij}^4 \right)^{1/4}$$

If $\max_{1 \leq i \leq n} \max_{1 \leq j \leq d} (\mathbb{E} \xi_{ij}^4)^{1/4} = O(1/\sqrt{n})$ as $n \rightarrow \infty$ as in the typical case in applications (where $\xi_{ij} = \frac{X_{ij}}{\sqrt{n}}$ for some X_{ij} not depending on n), this converges to 0 as long as $d/n \rightarrow 0$.

Remark 2.5. The inequality

$$\mathbb{E} |\xi_i|^4 \leq d \sum_{j=1}^d \mathbb{E} \xi_{ij}^4$$

can be obtained by applying Holder inequality

$$\sum_{k=1}^d |x_k y_k| \leq \left(\sum_{k=1}^d |x_k|^p \right)^{\frac{1}{p}} \left(\sum_{k=1}^d |y_k|^q \right)^{\frac{1}{q}}$$

with $p = q = 2$ and $x_k = \xi_{ik}^2$ and $y_k = 1$. In particular, we have

$$\begin{aligned} \mathbb{E} |\xi_i|^4 &= \mathbb{E} (\xi_{i1}^2 + \dots + \xi_{id}^2)^2 \\ &\leq \mathbb{E} [(\xi_{i1}^4 + \dots + \xi_{id}^4) (1 + \dots + 1)] \\ &= d \sum_{j=1}^d \mathbb{E} \xi_{ij}^4 \end{aligned}$$

Remark 2.6. Theorem 2.3 can be used to deduce Central Limit Theorems (CLTs) for $|W - a|^2$ under suitable conditions. For example, if $a = 0, \Sigma = I_d, \xi_i = X_i/\sqrt{n}$ for an i.i.d. sequence of random vectors $\{X_1, \dots, X_n\}$ with $\max_{1 \leq j \leq d} \mathbb{E} (X_{ij}^4) \leq C$, then by Theorem 2.3, Remark 2.4 and the CLT for chi-square random variables, we have, for $d \rightarrow \infty$ and $d = o(n)$

$$\frac{|W|^2 - d}{\sqrt{2d}} \rightarrow N(0, 1) \quad \text{in distribution.}$$

Proposition 2.2. Let X_1, \dots, X_n be i.i.d. standard Gaussian vectors in \mathbb{R}^d . Let $\{e_i\}_{i=1}^n$ be i.i.d. variables independent of $\{X_i\}_{i=1}^n$ with $\mathbb{E} e_1 = 0, \mathbb{E} e_1^2 = 1, \mathbb{E} e_1^4 < \infty$ and $\text{Var}(e_1^2) > 0$. Assume the law of e_1 does not depend on n . Set $W := n^{-1/2} \sum_{i=1}^n e_i X_i$ and let $Z \sim N(0, I_d)$. If

$$\sup_{x \geq 0} |\mathbb{P}(|W| \leq x) - \mathbb{P}(|Z| \leq x)| \rightarrow 0$$

as $d, n \rightarrow \infty$, we must have $d/n \rightarrow 0$.

Proof. Since $(|Z|^2 - d)/\sqrt{2d}$ converges in law to $N(0, 1)$ as $d \rightarrow \infty$, $(|W|^2 - d)/\sqrt{2d}$ also converges in law to $N(0, 1)$. Since W has the same law as $\sqrt{V} Z'$ by assumption, where $V := n^{-1} \sum_{i=1}^n e_i^2$ and $Z' \sim N(0, I_d)$ is independent of $\{e_i\}_{i=1}^\infty$, $(V|Z'|^2 - d)/\sqrt{2d}$ should also converge in law to $N(0, 1)$. Since

$$\frac{V|Z'|^2 - d}{\sqrt{2d}} = V \frac{|Z'|^2 - d}{\sqrt{2d}} + \sqrt{\frac{d}{2}}(V - 1) = (V - 1) \frac{|Z'|^2 - d}{\sqrt{2d}} + \frac{|Z'|^2 - d}{\sqrt{2d}} + \sqrt{\frac{d}{2}}(V - 1)$$

and the first term converges to 0 in probability,

$$\frac{|Z'|^2 - d}{\sqrt{2d}} + \sqrt{\frac{d}{2}}(V - 1)$$

must converge in law to $N(0, 1)$. In the above expression, the first term converges in law to $N(0, 1)$ and the first and second terms are independent, so this implies $\sqrt{d}(V - 1) = o_p(1)$ as $n \rightarrow \infty$. Since $\sqrt{n}(V - 1)$ converges in law to $N(0, \text{Var}(e_1^2))$, we must have $d/n \rightarrow 0$. \square

Remark 2.7. W in Proposition 2.2 can be regarded as a bootstrap approximation of Z . Remark 2.4 and Proposition 2.2 suggest that, in general, bootstrapping may not provide a more accurate approximation than the Gaussian approximation in terms of the dependence on dimension.

3 Application to Bootstrap Approximation on Balls

Notation 3.1. X_1, \dots, X_n : be a sequence of centered independent vectors in \mathbb{R}^d with finite fourth moments. $W := n^{-1/2} \sum_{i=1}^n X_i$ $\Sigma := \text{Var}(W)$ $Z \sim N(0, \Sigma)$ X_1^*, \dots, X_n^* : be i.i.d. draws from the empirical distribution of X

3.1 Empirical bootstrap approximation for $\mathbb{P}(W \in A)$

$W^* := \frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i^* - \bar{X})$, where $\bar{X} := \frac{1}{n} \sum_{i=1}^n X_i$ The bootstrap analog of 2.3 is given by:

Theorem 3.1. $\text{tr}(\Sigma^2) > 0$, for any $K > 0$, we have

$$\mathbb{P} \left(\sup_{A \in \mathcal{B}} |\mathbb{P}(W^* \in A | X) - \mathbb{P}(Z \in A)| > K \sqrt{\Delta_n} \right) \leq \frac{C}{K^2} \quad (7)$$

where

$$\Delta_n := \frac{1}{n \text{tr}(\Sigma^2)^{1/2}} \left(\sqrt{\sum_{i=1}^n \mathbb{E} |X_i|^4} + \sum_{j=1}^d \sqrt{\sum_{i=1}^n \mathbb{E} [X_{ij}^4]} \right)$$

Remark 3.1. Compared to the non-asymptotic bound for the quantity of $\sup_{A \in \mathcal{B}} |\mathbb{P}(W^* \in A | X) - \mathbb{P}(Z \in A)|$ under additional distribution assumption on X_i . Ours Theorem 3.1 provides better dependence on the **dimension** d ($d = o(n)$ v.s. $d = o(n^{1/2})$), **at least when** $\Sigma = I_d$; our result allows Σ **to be singular**; it's possible to give a non-asymptotic version of 7 but an exponential concentration if we also assume X_i are sub-Gaussian.

3.2 Wild bootstrap approximation for $\mathbb{P}(W \in A)$

Let $\{e_i\}_{i=1}^n$ be i.i.d. variables independent of $\{X_i\}_{i=1}^n$ with $\mathbb{E}e_1 = 0, \mathbb{E}e_1^2 = 1, \mathbb{E}e_1^4 < \infty$. The $W^o := \frac{1}{\sqrt{n}} \sum_{i=1}^n e_i X_i$ is the wild bootstrap approximation of W with multiplier variables e_1, \dots, e_n .

Theorem 3.2. $\text{tr}(\Sigma^2) > 0$, for any $K > 0$, we have

$$\mathbb{P} \left(\sup_{A \in \mathcal{B}} |\mathbb{P}(W^o \in A | X) - \mathbb{P}(Z \in A)| > K(\mathbb{E}e_1^4)^{1/4} \sqrt{\Delta_n} \right) \leq \frac{C}{K^2} \quad (8)$$

where Δ_n is defined in 3.1

Remark 3.2. Compared to the non-asymptotic bound for the quantity of $\sup_{A \in \mathcal{B}} |\mathbb{P}(W^o \in A | X) - \mathbb{P}(Z \in A)|$ under additional distribution assumption on X_i . Ours 3.2 provides better dependence on the n and $d(O(d/n)^{1/4}$ v.s. $O(d^2/n)^{1/5})$; ours does not require the **unit skewness assumption** $\mathbb{E}e_1^3 = 1$ on the multiplier variables; it's possible to give a non-asymptotic version of our result of 3.2.

4 Main Idea behind the Proof

4.1 Basic Decomposition

The proofs for Theorems 2.1 and 2.2-2.3 start with approximating the indicator function 1_A for $A \in \mathcal{A}$ or $A \in \mathcal{B}$ by an appropriate smooth function h . Then, the problem amounts to establishing an appropriate bound for $\mathbb{E}h(W) - \mathbb{E}h(Z)$.

To accomplish this, we will make use of a decomposition of $\mathbb{E}h(W) - \mathbb{E}h(Z)$ derived from the exchangeable pair approach in Stein's method for multivariate normal approximation by Chatterjee and Meckes (2008) and Reinert and Röllin (2009) along with a symmetry argument by Fang and Koike (2020a,b).

Lemma 4.1. Given a twice differentiable function $h : \mathbb{R}^d \rightarrow \mathbb{R}$ with bounded partial derivatives, we consider the Stein equation

$$\langle \text{Hess } f(w), \Sigma \rangle_{H.S.} - w \cdot \nabla f(w) = h(w) - \mathbb{E}h(Z), \quad w \in \mathbb{R}^d \quad (9)$$

then

$$f(w) = \int_0^1 -\frac{1}{2(1-s)} \int_{\mathbb{R}^d} \left[h \left(\sqrt{1-s}w + \sqrt{s}\Sigma^{1/2}z \right) - \mathbb{E}h(Z) \right] \phi(z) dz ds \quad (10)$$

is a solution to (9).

In the following we assume that f is thrice differentiable with bounded partial derivatives. This is true if Σ is invertible or h is thrice differentiable with bounded partial derivatives. Now we introduce the basic decomposition we will use later in proof.

Let $\{\xi'_1, \dots, \xi'_n\}$ be an independent copy of $\{\xi_1, \dots, \xi_n\}$, and let I be a random index uniformly chosen from $\{1, \dots, n\}$ and independent of $\{\xi_1, \dots, \xi_n, \xi'_1, \dots, \xi'_n\}$. Define $W' = W - \xi_I + \xi'_I$. It is easy to verify that (W, W') has the same distribution as (W', W) (exchangeability) and

$$\mathbb{E}(W' - W | W) = -\frac{W}{n} \quad (11)$$

From exchangeability and (11) we have, with $D = W' - W$

$$\begin{aligned} 0 &= \frac{n}{2} \mathbb{E}[D \cdot (\nabla f(W') + \nabla f(W))] \\ &= \mathbb{E}\left[\frac{n}{2} D \cdot (\nabla f(W') - \nabla f(W)) + nD \cdot \nabla f(W)\right] \\ &= \mathbb{E}\left[\frac{n}{2} \sum_{j,k=1}^d D_j D_k \partial_{jk} f(W) + R_2 + nD \cdot \nabla f(W)\right] \\ &= \mathbb{E}[\langle \text{Hess } f(W), \Sigma \rangle_{H.S.} - R_1 + R_2 - W \cdot \nabla f(W)] \end{aligned} \quad (12)$$

where

$$R_1 = \sum_{j,k=1}^d \mathbb{E}\left\{\left(\Sigma_{jk} - \frac{n}{2} D_j D_k\right) \partial_{jk} f(W)\right\} \quad (13)$$

and

$$R_2 = \frac{n}{2} \sum_{j,k,l=1}^d \mathbb{E} D_j D_k D_l U \partial_{jkl} f(W + (1-U)D) \quad (14)$$

and U is a uniform random variable on $[0, 1]$ independent of everything else. From (9) and (12) we have

$$\mathbb{E}h(W) - \mathbb{E}h(Z) = R_1 - R_2 \quad (15)$$

We further rewrite R_1 and R_2 respectively as follows. First, set

$$V = (V_{jk})_{1 \leq j, k \leq d} := \left(\mathbb{E} \left[\Sigma_{jk} - \frac{n}{2} D_j D_k \mid \xi \right] \right)_{1 \leq j, k \leq d}$$

Then we evidently have

$$R_1 = \sum_{j,k=1}^d \mathbb{E} V_{jk} \partial_{jk} f(W) = \mathbb{E} \langle V, \text{Hess } f(W) \rangle_{H.S.} \quad (16)$$

Also, one can easily verify that (cf. Eq.(22) of Chernozhukov, Chetverikov and Kato (2014))

$$V = \Sigma - \frac{1}{2} \sum_{i=1}^n \mathbb{E} [\xi_i \xi_i^\top] - \frac{1}{2} \sum_{i=1}^n \xi_i \xi_i^\top = (\Sigma - \text{Var}(W)) - \frac{1}{2} \sum_{i=1}^n (\xi_i \xi_i^\top - \mathbb{E} [\xi_i \xi_i^\top]) \quad (17)$$

Next, by exchangeability we have

$$\begin{aligned} & \mathbb{E}[D_j D_k D_l U \partial_{jkl} f(W + (1 - U)D)] \\ &= -\mathbb{E}[D_j D_k D_l U \partial_{jkl} f(W' - (1 - U)D)] \\ &= -\mathbb{E}[D_j D_k D_l U \partial_{jkl} f(W + UD)] \end{aligned} \quad (18)$$

Hence we obtain

$$R_2 = \frac{n}{4} \sum_{j,k,l=1}^d \mathbb{E}[D_j D_k D_l U \{\partial_{jkl} f(W + (1 - U)D) - \partial_{jkl} f(W + UD)\}] \quad (19)$$

4.2 Proof of Theorem 2.1

Since $\Sigma^{-1/2}W = \sum_{i=1}^n \Sigma^{-1/2}\xi_i$ and $\{\Sigma^{-1/2}x : x \in A\} \in \mathcal{A}$ for all $A \in \mathcal{A}$, it suffices to consider the case $\Sigma = I_d$. The proof is a combination of Bentkus (2003)'s smoothing, the decomposition (15), and a recursive argument by Raić (2019a). Fix $\beta_0 > 0$. Define

$$K(\beta_0) = \sup_W \frac{\sup_{A \in \mathcal{A}} |\mathbb{P}(W \in A) - \mathbb{P}(Z \in A)|}{\max \left\{ \beta_0, \left(\sum_{i \in \mathcal{I}} \mathbb{E}|\xi_i|^4 \right)^{1/2} \left(\left| \log \left(\sum_{i \in \mathcal{I}} \mathbb{E}|\xi_i|^4 \right) \right| \vee 1 \right) \right\}} \quad (20)$$

where the first supremum is taken over the family of all sums $W = \sum_{i \in \mathcal{I}} \xi_i$ of finite number of independent mean-zero random vectors with $\mathbb{E}|\xi_i|^4 < \infty$ and $\text{Var}(W) = I_d$. We will obtain a recursive inequality for $K(\beta_0)$ and prove that

$$K(\beta_0) \leq Cd^{1/4} \quad (21)$$

for an absolute constant C that does not depend on β_0 . Equation (4) then follows by sending $\beta_0 \rightarrow 0$.

Now we fix a $W = \sum_{i=1}^n \xi_i, n \geq 1$, in the aforementioned family

$$\bar{\beta} = \max \left\{ \beta_0, \left(\sum_{i=1}^n \mathbb{E}|\xi_i|^4 \right)^{1/2} \left(\left| \log \left(\sum_{i=1}^n \mathbb{E}|\xi_i|^4 \right) \right| \vee 1 \right) \right\}. \quad (22)$$

and for $A \in \mathcal{A}, \varepsilon > 0$, define

$$A^\varepsilon = \left\{ x \in \mathbb{R}^d : \text{dist}(x, A) \leq \varepsilon \right\} \quad \text{dist}(x, A) = \inf_{y \in A} |x - y|$$

Lemma 4.2. For any $A \in \mathcal{A}$ and $\varepsilon > 0$, there exists a function $h_{A,\varepsilon}$ (which depends only on A and ε) such that

$$h_{A,\varepsilon}(x) = 1 \text{ for } x \in A, \quad h_{A,\varepsilon}(x) = 0 \text{ for } x \in \mathbb{R}^d \setminus A^\varepsilon, \quad 0 \leq h_{A,\varepsilon}(x) \leq 1$$

and

$$M_1(h_{A,\varepsilon}) \leq \frac{C}{\varepsilon}, \quad M_2(h_{A,\varepsilon}) \leq \frac{C}{\varepsilon^2} \quad (23)$$

where C is an absolute constant that does not depend on A and ε .

Lemma 4.3. Let ϕ be the standard Gaussian density on \mathbb{R}^d , $d \geq 2$, and let A be a convex set in \mathbb{R}^d . Then

$$\int_{\partial A} \phi \leq 4d^{1/4} \quad (24)$$

Lemma 4.4. For any d -dimensional random vector W and any $\varepsilon > 0$,

$$\sup_{A \in \mathcal{A}} |P(W \in A) - P(Z \in A)| \leq 4d^{1/4}\varepsilon + \sup_{A \in \mathcal{A}} |\mathbb{E}h_{A,\varepsilon}(W) - \mathbb{E}h_{A,\varepsilon}(Z)| \quad (25)$$

where $h_{A,\varepsilon}$ is as in Lemma 4.2.

We now fix $A \in \mathcal{A}$ (will take sup later), $0 < \varepsilon \leq 1$, write $h := h_{A,\varepsilon}$ and proceed to bound $|\mathbb{E}h(W) - \mathbb{E}h(Z)|$ by the decomposition (15). Consider the solution f to the Stein equation (9) with $\Sigma = I_d$

$$f(w) = \int_0^1 -\frac{1}{2(1-s)} \int_{\mathbb{R}^d} [h(\sqrt{1-s}w + \sqrt{s}z) - \mathbb{E}h(Z)] \phi(z) dz ds$$

Since h has bounded partial derivatives up to the second order and $\Sigma = I_d$ is invertible, f is thrice differentiable with bounded partial derivatives. Using the integration by parts formula, we have for $1 \leq j, k, l \leq d$ and any constant $0 \leq c_0 \leq 1$ that

$$\begin{aligned} \partial_{jkl} f(w) &= \int_0^{c_0} \frac{1}{2\sqrt{s}} \int_{\mathbb{R}^d} \partial_j h(\sqrt{1-s}w + \sqrt{s}z) \partial_k \phi(z) dz ds \\ &\quad + \int_{c_0}^1 -\frac{1}{2s} \int_{\mathbb{R}^d} h(\sqrt{1-s}w + \sqrt{s}z) \partial_{jk} \phi(z) dz ds \end{aligned} \quad (26)$$

and

$$\begin{aligned} \partial_{jkl} f(w) &= \int_0^{c_0} \frac{\sqrt{1-s}}{2\sqrt{s}} \int_{\mathbb{R}^d} \partial_{jk} h(\sqrt{1-s}w + \sqrt{s}z) \partial_l \phi(z) dz ds \\ &\quad + \int_{c_0}^1 -\frac{\sqrt{1-s}}{2s} \int_{\mathbb{R}^d} \partial_j h(\sqrt{1-s}w + \sqrt{s}z) \partial_{kl} \phi(z) dz ds \end{aligned} \quad (27)$$

We first bound R_1 in (16). We will utilize the following lemma.

Lemma 4.5. For $k \geq 1$ and each map $a : \{1, \dots, d\}^k \rightarrow \mathbb{R}$, we have

$$\int_{\mathbb{R}^d} \left(\sum_{i_1, \dots, i_k=1}^d a(i_1, \dots, i_k) \frac{\partial_{i_1 \dots i_k} \phi(z)}{\phi(z)} \right)^2 \phi(z) dz \leq k! \sum_{i_1, \dots, i_k=1}^d (a(i_1, \dots, i_k))^2 \quad (28)$$

Now, using the expression of $\partial_{jk}f$ in (26) with $c_0 = \varepsilon^2$, we have

$$R_1 = R_{11} + R_{12}$$

where

$$R_{11} = \sum_{j,k=1}^d \mathbb{E} \left[V_{jk} \int_0^{\varepsilon^2} \frac{1}{2\sqrt{s}} \int_{\mathbb{R}^d} \partial_j h(\sqrt{1-s}W + \sqrt{s}z) \partial_k \phi(z) dz ds \right]$$

and

$$R_{12} = \sum_{j,k=1}^d \mathbb{E} \left[V_{jk} \int_{\varepsilon^2}^1 -\frac{1}{2s} \int_{\mathbb{R}^d} h(\sqrt{1-s}W + \sqrt{s}z) \partial_{jk} \phi(z) dz ds \right]$$

For R_{11} , we use the Cauchy-Schwarz inequality and the bounds (23) and (28) and obtain

$$\begin{aligned} |R_{11}| &= \left| \int_0^{\varepsilon^2} \frac{1}{2\sqrt{s}} \int_{\mathbb{R}^d} \mathbb{E} \sum_{j=1}^d \partial_j h(\sqrt{1-s}W + \sqrt{s}z) \sum_{k=1}^d V_{jk} \frac{\partial_k \phi(z)}{\phi(z)} \phi(z) dz ds \right| \\ &\leq \frac{C}{\varepsilon} \int_0^{\varepsilon^2} \frac{1}{2\sqrt{s}} \int_{\mathbb{R}^d} \mathbb{E} \left\{ \sum_{j=1}^d \left(\sum_{k=1}^d V_{jk} \frac{\partial_k \phi(z)}{\phi(z)} \right)^2 \right\}^{1/2} \phi(z) dz ds \\ &\leq \frac{C}{\varepsilon} \int_0^{\varepsilon^2} \frac{1}{2\sqrt{s}} \left\{ \int_{\mathbb{R}^d} \mathbb{E} \sum_{j=1}^d \left(\sum_{k=1}^d V_{jk} \frac{\partial_k \phi(z)}{\phi(z)} \right)^2 \phi(z) dz \right\}^{1/2} ds \\ &\leq \frac{C}{\varepsilon} \int_0^{\varepsilon^2} \frac{1}{2\sqrt{s}} \left\{ \mathbb{E} \sum_{j=1}^d \sum_{k=1}^d V_{jk}^2 \right\}^{1/2} ds \leq C \left\{ \sum_{j,k=1}^d \mathbb{E} V_{jk}^2 \right\}^{1/2} \end{aligned} \tag{29}$$

Recall that $\text{Var}(W) = \Sigma$

$$V = \Sigma - \frac{1}{2} \sum_{i=1}^n \mathbb{E} [\xi_i \xi_i^\top] - \frac{1}{2} \sum_{i=1}^n \xi_i \xi_i^\top = (\Sigma - \text{Var}(W)) - \frac{1}{2} \sum_{i=1}^n (\xi_i \xi_i^\top - \mathbb{E} [\xi_i \xi_i^\top])$$

we have

$$\mathbb{E} V_{jk}^2 = \frac{1}{4} \text{Var} \left[\sum_{i=1}^n \xi_{ij} \xi_{ik} \right] = \frac{1}{4} \sum_{i=1}^n \text{Var} [\xi_{ij} \xi_{ik}] \leq \frac{1}{4} \sum_{i=1}^n \mathbb{E} [\xi_{ij}^2 \xi_{ik}^2]$$

and therefore,

$$|R_{11}| \leq C \left\{ \sum_{j,k=1}^d \sum_{i=1}^n \mathbb{E} [\xi_{ij}^2 \xi_{ik}^2] \right\}^{1/2} = C \left\{ \sum_{i=1}^n \mathbb{E} \left[\sum_{j=1}^d \xi_{ij}^2 \right]^2 \right\}^{1/2} = C \left(\sum_{i=1}^n \mathbb{E} |\xi_i|^4 \right)^{1/2}$$

Applying similar arguments, we have, for R_{12} ,

$$\begin{aligned}
|R_{12}| &= \left| \int_{\varepsilon^2}^1 \left(-\frac{1}{2s} \right) \int_{\mathbb{R}^d} \mathbb{E} h(\sqrt{1-s}W + \sqrt{s}z) \sum_{j,k=1}^d V_{jk} \frac{\partial_{jk}\phi(z)}{\phi(z)} \phi(z) dz ds \right| \\
&\leq \int_{\varepsilon^2}^1 \frac{1}{2s} \int_{\mathbb{R}^d} \mathbb{E} \left| \sum_{j,k=1}^d V_{jk} \frac{\partial_{jk}\phi(z)}{\phi(z)} \right| \phi(z) dz ds \\
&\leq \int_{\varepsilon^2}^1 \frac{1}{2s} \left\{ \int_{\mathbb{R}^d} \mathbb{E} \left[\sum_{j,k=1}^d V_{jk} \frac{\partial_{jk}\phi(z)}{\phi(z)} \right]^2 \phi(z) dz \right\}^{1/2} ds \\
&\leq C \int_{\varepsilon^2}^1 \frac{1}{2s} \left\{ \mathbb{E} \sum_{j,k=1}^d V_{jk}^2 \right\}^{1/2} ds \leq C |\log \varepsilon| \left(\sum_{i=1}^n \mathbb{E} |\xi_i|^4 \right)^{1/2}
\end{aligned} \tag{30}$$

therefore,

$$|R_1| \leq C(|\log \varepsilon| \vee 1) \left(\sum_{i=1}^n \mathbb{E} |\xi_i|^4 \right)^{1/2} \tag{31}$$

Next, we bound R_2 . Take $0 < \eta \leq 1$ arbitrarily. Using the expression of $\partial_{jkl}f$ in (27) with $c_0 = \eta^2$ and the two equivalent expressions (14) and (19) for R_2 , we have

$$R_2 = R_{21} + R_{22}$$

where

$$\begin{aligned}
R_{21} &= \frac{1}{2} \sum_{i=1}^n \sum_{j,k,l=1}^d \mathbb{E} U (\xi'_{ij} - \xi_{ij}) (\xi'_{ik} - \xi_{ik}) (\xi'_{il} - \xi_{il}) \int_0^{\eta^2} \frac{\sqrt{1-s}}{2\sqrt{s}} \\
&\quad \times \int_{\mathbb{R}^d} \partial_{jk} h(\sqrt{1-s}(W + (1-U)(\xi'_i - \xi_i)) + \sqrt{s}z) \partial_l \phi(z) dz ds
\end{aligned}$$

and

$$\begin{aligned}
R_{22} &= \frac{1}{4} \sum_{i=1}^n \sum_{j,k,l=1}^d \mathbb{E} U (\xi'_{ij} - \xi_{ij}) (\xi'_{ik} - \xi_{ik}) (\xi'_{il} - \xi_{il}) \int_{\eta^2}^1 -\frac{\sqrt{1-s}}{2s} \\
&\quad \times \int_{\mathbb{R}^d} [\partial_j h(\sqrt{1-s}(W + (1-U)(\xi'_i - \xi_i)) + \sqrt{s}z) - \partial_j h(\sqrt{1-s}(W + U(\xi'_i - \xi_i)) + \sqrt{s}z)] \partial_{kl} \phi(z) dz ds \\
&= \frac{1}{4} \sum_{i=1}^n \sum_{j,k,l,m=1}^d \mathbb{E} U(1-2U) (\xi'_{ij} - \xi_{ij}) (\xi'_{ik} - \xi_{ik}) (\xi'_{il} - \xi_{il}) (\xi'_{im} - \xi_{im}) \int_{\eta^2}^1 -\frac{1-s}{2s} \\
&\quad \times \int_{\mathbb{R}^d} \partial_{jm} h(\sqrt{1-s}(W + (U + (1-2U)U')(\xi'_i - \xi_i)) + \sqrt{s}z) \partial_{kl} \phi(z) dz ds
\end{aligned} \tag{32}$$

where U' is a uniform random variable on $[0, 1]$ independent of everything else and we used the mean value theorem in the last equality. Let $W^{(i)} = W - \xi_i$ for $i \in \{1, \dots, n\}$. We will use the

fact that ∇h is non-zero only in $A^\varepsilon \setminus A$ and bound

$$\mathbb{P} \left(\sqrt{1-s} W^{(i)} \in A_i^\varepsilon \setminus A_i \mid U, U', \xi_i, \xi'_i \right)$$

where $0 < s < 1$ and A_i is a convex set which may depend on U, U', ξ_i, ξ'_i, s and z . Let Σ_i be the covariance matrix of $W^{(i)}$ and let σ_i be its smallest eigenvalue, which will be assumed to be positive in Case 1 below. We have

$$\begin{aligned} & \mathbb{P} \left(\sqrt{1-s} W^{(i)} \in A_i^\varepsilon \setminus A_i \mid U, U', \xi_i, \xi'_i \right) \\ &= \mathbb{P} \left(\Sigma_i^{-1/2} W^{(i)} \in \frac{1}{\sqrt{1-s}} \Sigma_i^{-1/2} (A_i^\varepsilon \setminus A_i) \mid U, U', \xi_i, \xi'_i \right) \\ &\leq 4d^{1/4} \frac{\varepsilon}{\sigma_i \sqrt{1-s}} + 2 \sup_{A \in \mathcal{A}} \left| \mathbb{P} \left(\Sigma_i^{-1/2} W^{(i)} \in A \right) - P(Z \in A) \right| \end{aligned} \quad (33)$$

where we used the $4d^{1/4}$ upper bound for the Gaussian surface area of any convex set in Lemma 4.3. From the definition in (20), we have

$$\begin{aligned} & \sup_{A \in \mathcal{A}} \left| \mathbb{P} \left(\Sigma_i^{-1/2} W^{(i)} \in A \right) - P(Z \in A) \right| \\ &\leq K(\beta_0) \max \left\{ \beta_0, \left(\sum_{j=1, j \neq i}^n \mathbb{E} |\Sigma_i^{-1/2} \xi_j|^4 \right)^{1/2} \left(\left| \log \left(\sum_{j=1, j \neq i}^n \mathbb{E} |\Sigma_i^{-1/2} \xi_j|^4 \right) \right| \vee 1 \right) \right\} \end{aligned} \quad (34)$$

Set $\beta_* = 0.19$ and $\sigma_* = (1 - \beta_*)^{1/2} = 0.9$. Recall that

$$\bar{\beta} = \max \left\{ \beta_0, \left(\sum_{i=1}^n \mathbb{E} |\xi_i|^4 \right)^{1/2} \left(\left| \log \left(\sum_{i=1}^n \mathbb{E} |\xi_i|^4 \right) \right| \vee 1 \right) \right\}$$

- Case-1: $\bar{\beta} \leq \beta_*/d^{1/4}$.

In this case, we have

$$\mathbb{E} |\xi_i|^2 \leq \sqrt{\mathbb{E} |\xi_i|^4} \leq \bar{\beta} \leq \beta_*/d^{1/4} \leq \beta_*$$

for each unit vector $u \in \mathbb{R}^d$

$$\langle \Sigma_i u, u \rangle = u^\top \Sigma_i u = u^\top (I_d - \mathbb{E} \xi_i \xi_i^\top) u = 1 - \mathbb{E} (\xi_i \cdot u)^2 \geq 1 - \mathbb{E} |\xi_i|^2 \geq 1 - \beta_*$$

this implies $\sigma_i \geq \sigma_*$. Note that $x^{1/2}(|\log x| \vee 1)$ is an increasing function. Therefore, from (34), we have, by increasing $\sum_{j=1, j \neq i}^n \mathbb{E} |\Sigma_i^{-1/2} \xi_j|^4$ to $\frac{1}{\sigma_*^4} \sum_{j=1}^n \mathbb{E} |\xi_j|^4$,

$$\sup_{A \in \mathcal{A}} \left| \mathbb{P} \left(\Sigma_i^{-1/2} W^{(i)} \in A \right) - P(Z \in A) \right| \leq K(\beta_0) \max \left\{ \beta_0, \frac{2\bar{\beta}}{\sigma_*^2} \right\} \leq CK(\beta_0) \bar{\beta} \quad (35)$$

Applying (23), (33), (35) and (3) we have

$$|R_{21}| \leq \frac{C}{\varepsilon^2} \sum_{i=1}^n \mathbb{E} |\xi_i|^3 \left(d^{1/4} \varepsilon + K(\beta_0) \bar{\beta} \right) \eta \quad (36)$$

and

$$|R_{22}| \leq \frac{C}{\varepsilon^2} \sum_{i=1}^n \mathbb{E} |\xi_i|^4 \left(d^{1/4} \varepsilon + K(\beta_0) \bar{\beta} \right) |\log \eta| \quad (37)$$

Now, if $\sum_{i=1}^n \mathbb{E} |\xi_i|^4 < \sum_{i=1}^n \mathbb{E} |\xi_i|^3$, choose $\eta = \sum_{i=1}^n \mathbb{E} |\xi_i|^4 / \sum_{i=1}^n \mathbb{E} |\xi_i|^3 < 1$. Note that we have by the Cauchy-Schwarz inequality

$$\sum_{i=1}^n \mathbb{E} |\xi_i|^3 \leq \sqrt{\sum_{i=1}^n \mathbb{E} |\xi_i|^2 \sum_{i=1}^n \mathbb{E} |\xi_i|^4} = \sqrt{d \sum_{i=1}^n \mathbb{E} |\xi_i|^4}$$

Thus we obtain

$$|\log \eta| \leq \frac{1}{2} \log d - \frac{1}{2} \log \left(\sum_{i=1}^n \mathbb{E} |\xi_i|^4 \right)$$

Since $\left(\sum_{i=1}^n \mathbb{E} |\xi_i|^4 \right)^{1/2} \leq \bar{\beta} \leq \beta_* / d^{1/4}$, we have $\frac{1}{2} \log d \leq 2 \log \beta_* - \log \left(\sum_{i=1}^n \mathbb{E} |\xi_i|^4 \right)$ and

$$|\log \eta| \leq C \left| \log \left(\sum_{i=1}^n \mathbb{E} |\xi_i|^4 \right) \right|$$

Hence, we have

$$|R_{21}| + |R_{22}| \leq \frac{C}{\varepsilon^2} \sum_{i=1}^n \mathbb{E} |\xi_i|^4 \left(d^{1/4} \varepsilon + K(\beta_0) \bar{\beta} \right) \left(\left| \log \left(\sum_{i=1}^n \mathbb{E} |\xi_i|^4 \right) \right| \vee 1 \right) \quad (38)$$

This inequality also holds true if $\sum_{i=1}^n \mathbb{E} |\xi_i|^4 \geq \sum_{i=1}^n \mathbb{E} |\xi_i|^3$ by taking $\eta = 1$.

$$\begin{aligned} & \sup_{A \in \mathcal{A}} |P(W \in A) - P(Z \in A)| \\ & \leq 4d^{1/4} \varepsilon + C(|\log \varepsilon| \vee 1) \left(\sum_{i=1}^n \mathbb{E} |\xi_i|^4 \right)^{1/2} \\ & \quad + \frac{C}{\varepsilon^2} \sum_{i=1}^n \mathbb{E} |\xi_i|^4 \left(d^{1/4} \varepsilon + K(\beta_0) \bar{\beta} \right) \left(\left| \log \left(\sum_{i=1}^n \mathbb{E} |\xi_i|^4 \right) \right| \vee 1 \right) \end{aligned} \quad (39)$$

Choose $\varepsilon = \min \left\{ \left[2C \sum_{i=1}^n \mathbb{E} |\xi_i|^4 \left(\left| \log \left(\sum_{i=1}^n \mathbb{E} |\xi_i|^4 \right) \right| \vee 1 \right) \right]^{1/2}, 1 \right\}$ with the same absolute constant C as in the third term on the right-hand side of (39). If $\varepsilon < 1$, then (39) can be simplified to

$$\sup_{A \in \mathcal{A}} |P(W \in A) - P(Z \in A)| \leq \left(Cd^{1/4} + \frac{K(\beta_0)}{2} \right) \bar{\beta}$$

hence

$$\frac{\sup_{A \in \mathcal{A}} |P(W \in A) - P(Z \in A)|}{\bar{\beta}} \leq Cd^{1/4} + \frac{K(\beta_0)}{2} \quad (40)$$

If $\varepsilon = 1$, then $\sum_{i=1}^n \mathbb{E} |\xi_i|^4$ and $\bar{\beta}$ are bounded away from 0 by an absolute constant; hence

$$\frac{\sup_{A \in \mathcal{A}} |P(W \in A) - P(Z \in A)|}{\bar{\beta}} \leq \frac{1}{\bar{\beta}} \leq C \quad (41)$$

- Case-2: $\bar{\beta} > \beta_* / d^{1/4}$, we trivially estimate

$$\frac{\sup_{A \in \mathcal{A}} |P(W \in A) - P(Z \in A)|}{\bar{\beta}} \leq \frac{1}{\bar{\beta}} \leq \frac{d^{1/4}}{\beta_*} \leq Cd^{1/4} \quad (42)$$

Combining both cases together, we have

$$\frac{\sup_{A \in \mathcal{A}} |P(W \in A) - P(Z \in A)|}{\bar{\beta}} \leq Cd^{1/4} + \frac{K(\beta_0)}{2}$$

Note that the right-hand side of the above bound does not depend on W . Taking supremum over W , we obtain

$$K(\beta_0) \leq Cd^{1/4} + \frac{K(\beta_0)}{2} \quad (43)$$

4.3 The Outline of the Proof for Theorem 2.2

We first note that, for any $d \times d$ orthogonal matrix U , we have $UW = \sum_{i=1}^n U\xi_i$, $UZ \sim N(0, U\Sigma U^\top)$, $\delta(UW, U\Sigma U^\top) = \delta(W, \Sigma)$ and $UB \in \mathcal{B}$ for all $B \in \mathcal{B}$. Therefore, it is enough to prove (5) when Σ is diagonal with positive entries. The proof is a combination of Zhilova (2020)'s smoothing, a Gaussian anti-concentration inequality for ellipsoids by Giessing and Fan (2020), the decomposition (15), and a recursive argument by Raič (2019a).

Fix $\beta_0 > 0$. Define

$$K'(\beta_0) = \sup_{W, \Sigma} \frac{\sup_{A \in \mathcal{B}} |\mathbb{P}(W \in A) - \mathbb{P}(\Sigma^{1/2} Z_0 \in A)|}{\max \{\beta_0, \Psi(\delta(W, \Sigma))\}} \quad (44)$$

where $Z_0 \sim N(0, I_d)$ and the first supremum is taken over the family of all sums $W = \sum_{i \in \mathcal{I}} \xi_i$ of finite number of independent centered random vectors with $\mathbb{E} |\xi_i|^4 < \infty$, and diagonal matrices Σ with positive entries. We will obtain a recursive inequality for $K'(\beta_0)$ and prove that

$$K'(\beta_0) \leq C \quad (45)$$

for an absolute constant C that does not depend on β_0 . Equation (5) then follows by sending $\beta_0 \rightarrow 0$.

Now we fix a $W = \sum_{i=1}^n \xi_i$, $n \geq 1$, and Σ in the aforementioned family (will take sup in (5.53)). Let

$$\bar{\beta} = \max \{ \beta_0, \Psi(\delta(W, \Sigma)) \} \quad (46)$$

We write σ_j for the j -th diagonal entry of $\Sigma^{1/2}$. To prove theorem 2.2, we need some technical lemmas that is applicable in the case of Euclidean ball.

Lemma 4.6. For any $A \in \mathcal{B}$ and $\varepsilon > 0$, there exists a C^∞ function $\tilde{h}_{A,\varepsilon}$ (which depends only on A and ε) such that

$$\tilde{h}_{A,\varepsilon}(x) = 1 \text{ for } x \in A, \quad \tilde{h}_{A,\varepsilon}(x) = 0 \text{ for } x \in \mathbb{R}^d \setminus A^\varepsilon, \quad 0 \leq \tilde{h}_{A,\varepsilon}(x) \leq 1 \quad (47)$$

and

$$M_r(\tilde{h}_{A,\varepsilon}) \leq \frac{C}{\varepsilon^r} \quad \text{for } r = 1, 2, 3, 4 \quad (48)$$

and

$$\sup_{x \in \mathbb{R}^d} \left| \left\langle M, \text{Hess } \tilde{h}_{A,\varepsilon}(x) \right\rangle_{H.S.} \right| \leq \frac{C}{\varepsilon^2} \left(\|M\|_{H.S.} + \sum_{j=1}^d |M_{jj}| \right) \quad (49)$$

for any $d \times d$ matrix $M = (M_{jk})_{1 \leq j, k \leq d}$, where C is an absolute constant that does not depend on A, ε or M .

Lemma 4.7. For any d -dimensional random vector W and any $\varepsilon > 0$,

$$\sup_{A \in \mathcal{B}} |P(W \in A) - P(Z \in A)| \leq \sup_{A \in \mathcal{B}} \mathbb{P}(Z \in A^\varepsilon \setminus A) + \sup_{A \in \mathcal{B}} \left| \mathbb{E} \tilde{h}_{A,\varepsilon}(W) - \mathbb{E} \tilde{h}_{A,\varepsilon}(Z) \right| \quad (50)$$

where $\tilde{h}_{A,\varepsilon}$ is as in Lemma 4.6.

Set $\tilde{\sigma} := \text{tr}(\Sigma^2)^{1/4}$. The following anti-concentration inequality is an immediate consequence of Giessing and Fan (2020, Corollary 5):

Lemma 4.8. Assume $\tilde{\sigma} > 0$. For any $\varepsilon > 0$,

$$\sup_{A \in \mathcal{B}} \mathbb{P}(Z \in A^\varepsilon \setminus A) \leq C \tilde{\sigma}^{-1} \varepsilon$$

Lemma 4.9. Ψ is an increasing function on $(0, \infty)$. Moreover, $\Psi(cx) \leq (c + \Psi(c))\Psi(x)$ for all $x > 0$ and $c \geq 1$.

We now fix $A \in \mathcal{B}$ (will take sup in (5.51)), $0 < \varepsilon \leq \tilde{\sigma}$, write $h := \tilde{h}_{A,\varepsilon}$ and proceed to bound $|\mathbb{E}h(W) - \mathbb{E}h(Z)|$ by the decomposition (15). Consider the solution f to the Stein equation (9), which is given by (10)

$$f(w) = \int_0^1 -\frac{1}{2(1-s)} \int_{\mathbb{R}^d} [h(\sqrt{1-s}w + \sqrt{s}\Sigma^{1/2}z) - \mathbb{E}h(Z)] \phi(z) dz ds$$

Since h has bounded partial derivatives up to the third order, f is thrice differentiable with bounded partial derivatives. Using the integration by parts formula, we have for $1 \leq j, k, l \leq d$ and any $0 \leq c_0 \leq 1$ that

$$\begin{aligned} \partial_{jk} f(w) &= \int_0^{c_0} \frac{1}{2\sqrt{s}} \int_{\mathbb{R}^d} \partial_j h(\sqrt{1-s}w + \sqrt{s}\Sigma^{1/2}z) \sigma_k^{-1} \partial_k \phi(z) dz ds \\ &\quad + \int_{c_0}^1 -\frac{1}{2s} \int_{\mathbb{R}^d} h(\sqrt{1-s}w + \sqrt{s}\Sigma^{1/2}z) \sigma_j^{-1} \sigma_k^{-1} \partial_{jk} \phi(z) dz ds \end{aligned} \tag{51}$$

and

$$\begin{aligned} \partial_{jkl} f(w) &= \int_0^{c_0} \frac{\sqrt{1-s}}{2\sqrt{s}} \int_{\mathbb{R}^d} \partial_{jk} h(\sqrt{1-s}w + \sqrt{s}\Sigma^{1/2}z) \sigma_l^{-1} \partial_l \phi(z) dz ds \\ &\quad + \int_{c_0}^1 \frac{\sqrt{1-s}}{2s^{3/2}} \int_{\mathbb{R}^d} h(\sqrt{1-s}w + \sqrt{s}\Sigma^{1/2}z) \sigma_j^{-1} \sigma_k^{-1} \sigma_l^{-1} \partial_{jkl} \phi(z) dz ds. \end{aligned} \tag{52}$$

We first bound R_1 in decomposition. Using the expression of $\partial_{jk} f$ in (51) with $c_0 = (\varepsilon/\tilde{\sigma})^2$, we have

$$R_1 = R_{11} + R_{12}$$

where

$$R_{11} = \sum_{j,k=1}^d \mathbb{E} \left[V_{jk} \int_0^{(\varepsilon/\tilde{\sigma})^2} \frac{1}{2\sqrt{s}} \int_{\mathbb{R}^d} \partial_j h(\sqrt{1-s}W + \sqrt{s}\Sigma^{1/2}z) \sigma_k^{-1} \partial_k \phi(z) dz ds \right]$$

and

$$R_{12} = \sum_{j,k=1}^d \mathbb{E} \left[V_{jk} \int_{(\varepsilon/\tilde{\sigma})^2}^1 -\frac{1}{2s} \int_{\mathbb{R}^d} h(\sqrt{1-s}W + \sqrt{s}\Sigma^{1/2}z) \sigma_j^{-1} \sigma_k^{-1} \partial_{jk} \phi(z) dz ds \right]$$

For R_{11} , applying analogous arguments to (29), we obtain

$$|R_{11}| \leq C \tilde{\sigma}^{-1} \left\{ \sum_{j,k=1}^d \sigma_k^{-2} \mathbb{E} V_{jk}^2 \right\}^{1/2} \leq C \left\{ \sum_{j,k=1}^d (\sigma_j \sigma_k)^{-2} \mathbb{E} V_{jk}^2 \right\}^{1/2}$$

where we used the inequality $\tilde{\sigma} \geq \sigma_j$ to derive the last inequality. The triangle inequality yields, for V in (17),

$$\left\{ \sum_{j,k=1}^d (\sigma_j \sigma_k)^{-2} \mathbb{E} V_{jk}^2 \right\}^{1/2} \leq \|I_d - \text{Var}(\Sigma^{-1/2}W)\|_{H.S.} + \frac{1}{2} \left\{ \sum_{j,k=1}^d (\sigma_j \sigma_k)^{-2} \text{Var} \left[\sum_{i=1}^n \xi_{ij} \xi_{ik} \right] \right\}^{1/2}$$

Besides, we have

$$\begin{aligned}
\left\{ \sum_{j,k=1}^d (\sigma_j \sigma_k)^{-2} \operatorname{Var} \left[\sum_{i=1}^n \xi_{ij} \xi_{ik} \right] \right\}^{1/2} &= \left\{ \sum_{i=1}^n \sum_{j,k=1}^d (\sigma_j \sigma_k)^{-2} \operatorname{Var} [\xi_{ij} \xi_{ik}] \right\}^{1/2} \\
&\leq \left\{ \sum_{i=1}^n \sum_{j,k=1}^d (\sigma_j \sigma_k)^{-2} \mathbb{E} \xi_{ij}^2 \xi_{ik}^2 \right\}^{1/2} \\
&= \left\{ \sum_{i=1}^n \mathbb{E} \left[\left(\sum_{j=1}^d \sigma_j^{-2} \xi_{ij}^2 \right)^2 \right] \right\}^{1/2} \\
&= \left(\sum_{i=1}^n \mathbb{E} |\Sigma^{-1/2} \xi_i|^4 \right)^{1/2}
\end{aligned}$$

Consequently, we obtain

$$|R_{11}| \leq C \delta(W, \Sigma)$$

For R_{12} , we apply analogous arguments to (30) and obtain

$$|R_{12}| \leq C |\log(\varepsilon/\tilde{\sigma})| \left\{ \sum_{j,k=1}^d (\sigma_j \sigma_k)^{-2} \mathbb{E} V_{jk}^2 \right\}^{1/2} \leq C |\log(\varepsilon/\tilde{\sigma})| \delta(W, \Sigma)$$

Therefore,

$$|R_1| \leq C(|\log(\varepsilon/\tilde{\sigma})| \vee 1) \delta(W, \Sigma) \quad (53)$$

Next, we bound R_2 in 19. Using the expression of $\partial_{jkl} f$ in 52 with $c_0 = (\varepsilon/\tilde{\sigma})^2$, we have

$$\begin{aligned}
R_{21} &= \frac{1}{4} \sum_{i=1}^n \sum_{j,k,l=1}^d \mathbb{E} U (\xi'_{ij} - \xi_{ij}) (\xi'_{ik} - \xi_{ik}) (\xi'_{il} - \xi_{il}) \int_0^{(\varepsilon/\tilde{\sigma})^2} \frac{\sqrt{1-s}}{2\sqrt{s}} \times \int_{\mathbb{R}^d} \phi_{21}(z, s) dz ds \\
R_{22} &= \frac{1}{4} \sum_{i=1}^n \sum_{j,k,l=1}^d \mathbb{E} U (\xi'_{ij} - \xi_{ij}) (\xi'_{ik} - \xi_{ik}) (\xi'_{il} - \xi_{il}) \int_{(\varepsilon/\tilde{\sigma})^2}^1 \frac{\sqrt{1-s}}{2s^{3/2}} \times \int_{\mathbb{R}^d} \phi_{22}(z, s) dz ds
\end{aligned}$$

and

$$\begin{aligned}
\phi_{21}(z, s) &= \left[\partial_{jk} h \left(\sqrt{1-s} (W + (1-U)(\xi'_i - \xi_i)) + \sqrt{s} \Sigma^{1/2} z \right) \right. \\
&\quad \left. - \partial_{jk} h \left(\sqrt{1-s} (W + U(\xi'_i - \xi_i)) + \sqrt{s} \Sigma^{1/2} z \right) \right] \sigma_l^{-1} \partial_l \phi(z) \\
\phi_{22}(z, s) &= \left[h \left(\sqrt{1-s} (W + (1-U)(\xi'_i - \xi_i)) + \sqrt{s} \Sigma^{1/2} z \right) \right. \\
&\quad \left. - h \left(\sqrt{1-s} (W + U(\xi'_i - \xi_i)) + \sqrt{s} \Sigma^{1/2} z \right) \right] \sigma_j^{-1} \sigma_k^{-1} \sigma_l^{-1} \partial_{jkl} \phi(z)
\end{aligned}$$

Let $W^{(i)} = W - \xi_i$ for $i \in \{1, \dots, n\}$. We will use the mean value theorem for the differences involving h in the above two expressions as in (5.25), the fact that ∇h is non-zero only in $A^\varepsilon \setminus A$ and bound

$$\mathbb{P} \left(\sqrt{1-s} W^{(i)} \in A_i^\varepsilon \setminus A_i \mid U, U', \xi_i, \xi'_i \right)$$

where $0 < s < 1$, U' is a uniform random variable on $[0,1]$ independent of everything else, and A_i is a Euclidean ball which may depend on U, U', ξ_i, ξ'_i, s and $\Sigma^{1/2} z$. We have by Lemma 4.8

$$\mathbb{P} \left(\sqrt{1-s} W^{(i)} \in A_i^\varepsilon \setminus A_i \mid U, U', \xi_i, \xi'_i \right) \leq C \frac{\varepsilon}{\tilde{\sigma} \sqrt{1-s}} + 2 \sup_{A \in \mathcal{B}} \left| \mathbb{P} \left(W^{(i)} \in A \right) - \mathbb{P}(Z \in A) \right| \quad (54)$$

From (44), we have

$$\sup_{A \in \mathcal{B}} \left| \mathbb{P} \left(W^{(i)} \in A \right) - \mathbb{P}(Z \in A) \right| \leq K'(\beta_0) \max \left\{ \beta_0, \Psi \left(\delta \left(W^{(i)}, \Sigma \right) \right) \right\} \quad (55)$$

Since

$$\begin{aligned} \left\| \text{Var} \left(\Sigma^{-1/2} W \right) - \text{Var} \left(\Sigma^{-1/2} W^{(i)} \right) \right\|_{H.S.} &= \sqrt{\sum_{j,k=1}^d \left(\mathbb{E} \left(\Sigma^{-1/2} \xi_j \right)_j \left(\Sigma^{-1/2} \xi_k \right)_k \right)^2} \\ &\leq \mathbb{E} \left| \Sigma^{-1/2} \xi_i \right|^2 \leq \sqrt{\mathbb{E} \left| \Sigma^{-1/2} \xi_i \right|^4} \end{aligned}$$

and $\sqrt{x} + \sqrt{y} \leq \sqrt{2(x+y)}$ for any $x, y \geq 0$, we have

$$\begin{aligned} \delta \left(W^{(i)}, \Sigma \right) &\leq \left\| I_d - \text{Var} \left(\Sigma^{-1/2} W \right) \right\|_{H.S.} + \sqrt{\mathbb{E} \left| \Sigma^{-1/2} \xi_i \right|^4} + \sqrt{\sum_{\substack{j=1 \\ j \neq i}}^n \mathbb{E} \left| \Sigma^{-1/2} \xi_j \right|^4} \\ &\leq \left\| I_d - \text{Var} \left(\Sigma^{-1/2} W \right) \right\|_{H.S.} + \sqrt{2 \sum_{j=1}^n \mathbb{E} \left| \Sigma^{-1/2} \xi_j \right|^4} \leq \sqrt{2} \delta(W, \Sigma) \end{aligned}$$

Hence, we obtain by Lemma 4.9

$$\Psi \left(\delta \left(W^{(i)}, \Sigma \right) \right) \leq 2\sqrt{2} \Psi(\delta(W, \Sigma)) \leq 2\sqrt{2} \bar{\beta}$$

Thus we conclude

$$\sup_{A \in \mathcal{B}} \left| \mathbb{P} \left(W^{(i)} \in A \right) - P(Z \in A) \right| \leq K'(\beta_0) \max \left\{ \beta_0, 2\sqrt{2} \bar{\beta} \right\} = 2\sqrt{2} K'(\beta_0) \bar{\beta} \quad (56)$$

Using the mean value theorem for R_{21}, R_{22} and applying (48), (54), (56) and (3) we have

$$|R_{21}| + |R_{22}| \leq \frac{C \tilde{\sigma}^2}{\varepsilon^2} \sum_{i=1}^n \mathbb{E} \left| \Sigma^{-1/2} \xi_i \right|^4 \left(\frac{\varepsilon}{\tilde{\sigma}} + 2\sqrt{2} K'(\beta_0) \bar{\beta} \right) \quad (57)$$

where we also used the inequality $\max_{1 \leq j \leq d} \sigma_j \leq \tilde{\sigma}$. From Lemmas 4.7-4.8, Equations (15), (53) and (57) we have

$$\begin{aligned} & \sup_{A \in \mathcal{B}} |\mathbb{P}(W \in A) - \mathbb{P}(Z \in A)| \\ & \leq C\tilde{\sigma}^{-1}\varepsilon + C(|\log(\varepsilon/\tilde{\sigma})| \vee 1)\delta(W, \Sigma) + \frac{C\tilde{\sigma}^2}{\varepsilon^2} \sum_{i=1}^n \mathbb{E} \left| \Sigma^{-1/2} \xi_i \right|^4 \left(\frac{\varepsilon}{\tilde{\sigma}} + K'(\beta_0) \bar{\beta} \right) \end{aligned} \quad (58)$$

Choose $\varepsilon = \min \left\{ \tilde{\sigma} \left[2C \sum_{i=1}^n \mathbb{E} \left| \Sigma^{-1/2} \xi_i \right|^4 \right]^{1/2}, \tilde{\sigma} \right\}$ for the same absolute constant C as in the third term on the right-hand side of (58)

- If $\varepsilon < \tilde{\sigma}$, then from (58)

$$\sup_{A \in \mathcal{B}} |\mathbb{P}(W \in A) - \mathbb{P}(Z \in A)| \leq \left(C + \frac{K'(\beta_0)}{2} \right) \bar{\beta}$$

hence

$$\frac{\sup_{A \in \mathcal{B}} |\mathbb{P}(W \in A) - \mathbb{P}(Z \in A)|}{\bar{\beta}} \leq C + \frac{K'(\beta_0)}{2} \quad (59)$$

- If $\varepsilon = \tilde{\sigma}$, then $\sum_{i=1}^n \mathbb{E} \left| \Sigma^{-1/2} \xi_i \right|^4$ and $\bar{\beta}$ are bounded away from 0 by an absolute constant; hence,

$$\frac{\sup_{A \in \mathcal{A}} |\mathbb{P}(W \in A) - \mathbb{P}(Z \in A)|}{\bar{\beta}} \leq \frac{1}{\bar{\beta}} \leq C$$

Note that the right-hand sides of the above two bounds do not depend on W or Σ . Taking supremum over W and Σ , we obtain

$$K'(\beta_0) \leq C + \frac{K'(\beta_0)}{2} \quad (60)$$

This implies (45) hence (5).

4.4 Proof of Theorem 2.3

Fix $A \in \mathcal{B}$ (will take sup in the end of the proof), $\varepsilon > 0$, write $h := \tilde{h}_{A,\varepsilon}$ as in Lemma 4.6 and proceed to bound $|\mathbb{E}h(W) - \mathbb{E}h(Z)|$ by the decomposition (15). Consider the solution f to the Stein equation (9), which is given by (10). Note that we can rewrite f as

$$f(w) = \int_0^1 -\frac{1}{2(1-s)} \mathbb{E}[h(\sqrt{1-s}w + Z) - \mathbb{E}h(Z)] ds$$

Since h has bounded partial derivatives up to the fourth order, f is four times differentiable and

$$\nabla^r f(w) = \int_0^1 -\frac{(1-s)^{r/2-1}}{2} \mathbb{E} [\nabla^r h(\sqrt{1-s}w + \sqrt{s}Z)] ds \quad (61)$$

We first bound R_1 in (16). Using (61), we obtain

$$R_1 = -\frac{1}{2} \int_0^1 \mathbb{E} [\langle V, \text{Hess } h(W^s) \rangle_{H.S.}] ds$$

where $W^s := \sqrt{1-s}W + \sqrt{s}Z$. Since ∇h is non-zero only in $A^\varepsilon \setminus A$, we have

$$R_1 = -\frac{1}{2} \int_0^1 \mathbb{E} [\langle V, \text{Hess } h(W^s) \rangle_{H.S.} \cdot 1_{\{W^s \in A^\varepsilon \setminus A\}}] ds$$

Therefore, using (49) we obtain

$$\begin{aligned} |R_1| &\leq \frac{1}{2} \int_0^1 \mathbb{E} |\langle V, \text{Hess } h(W^s) \rangle_{H.S.}| \cdot 1_{\{W^s \in A^\varepsilon \setminus A\}} ds \\ &\leq \frac{C}{\varepsilon^2} \int_0^1 \mathbb{E} \left(\|V\|_{H.S.} + \sum_{j=1}^d |V_{jj}| \right) 1_{\{W^s \in A^\varepsilon \setminus A\}} ds \\ &= \frac{C}{\varepsilon^2} \int_0^1 \mathbb{E} \left(\|V\|_{H.S.} + \sum_{j=1}^d |V_{jj}| \right) \mathbb{P}(W^s \in A^\varepsilon \setminus A \mid \xi) ds \end{aligned}$$

Since Z is independent of ξ , Lemma 4.8 yields

$$\mathbb{P}(W^s \in A^\varepsilon \setminus A \mid \xi) \leq \frac{C\varepsilon}{\tilde{\sigma}\sqrt{s}}$$

Thus we deduce

$$|R_1| \leq \frac{C}{\tilde{\sigma}\varepsilon} \mathbb{E} \left(\|V\|_{H.S.} + \sum_{j=1}^d |V_{jj}| \right) \int_0^1 \frac{1}{\sqrt{s}} ds \leq \frac{C}{\tilde{\sigma}\varepsilon} \mathbb{E} \left(\|V\|_{H.S.} + \sum_{j=1}^d |V_{jj}| \right)$$

Using (17) and the triangle inequality, we obtain

$$\begin{aligned} &\mathbb{E} \left(\|V\|_{H.S.} + \sum_{j=1}^d |V_{jj}| \right) \\ &\leq \|\Sigma - \text{Var}(W)\|_{H.S.} + \sum_{j=1}^d |\Sigma_{jj} - \text{Var}(W_j)| + \frac{1}{2} \sqrt{\sum_{j,k=1}^d \sum_{i=1}^n \text{Var}(\xi_{ij}\xi_{ik})} + \frac{1}{2} \sum_{j=1}^d \sqrt{\sum_{i=1}^n \text{Var}(\xi_{ij}^2)} \\ &\leq \|\Sigma - \text{Var}(W)\|_{H.S.} + \sum_{j=1}^d |\Sigma_{jj} - \text{Var}(W_j)| + \frac{1}{2} \sqrt{\sum_{j,k=1}^d \sum_{i=1}^n \mathbb{E}[\xi_{ij}^2\xi_{ik}^2]} + \frac{1}{2} \sum_{j=1}^d \sqrt{\sum_{i=1}^n \mathbb{E}[\xi_{ij}^4]} \\ &\leq \tilde{\delta}(W, \Sigma) \end{aligned}$$

Therefore, we conclude

$$|R_1| \leq \frac{C}{\tilde{\sigma}\varepsilon} \tilde{\delta}(W, \Sigma) \tag{62}$$

Next we bound R_2 in (19). We rewrite it as

$$R_2 = \frac{n}{4} \sum_{j,k,l,m=1}^d \mathbb{E} [D_j D_k D_l D_m U (1 - 2U) \partial_{jklm} f(W + \tilde{D})] \quad (63)$$

where $\tilde{D} := UD + U'(1 - 2U)D$ and U' is a uniform random variable on $[0,1]$ independent of everything else. Now we set $\tilde{W}^s := \sqrt{1-s}(W + \tilde{D}) + \sqrt{s}Z$. Then, using (61), we can rewrite R_2 as

$$R_2 = n \int_0^1 -\frac{1-s}{8} \mathbb{E} U(1-2U) \left\langle \nabla^4 h(\tilde{W}^s), D^{\otimes 4} \right\rangle ds$$

Since ∇h is non-zero only in $A^\varepsilon \setminus A$, we can further rewrite it as

$$R_2 = n \int_0^1 -\frac{1-s}{8} \mathbb{E} U(1-2U) \left\langle \nabla^4 h(\tilde{W}^s), D^{\otimes 4} \right\rangle 1_{\{\tilde{W}^s \in A^\varepsilon \setminus A\}} ds$$

Therefore, using (48), we obtain

$$\begin{aligned} |R_2| &\leq \frac{n}{8} \int_0^1 \mathbb{E} \left| \left\langle \nabla^4 h(\tilde{W}^s), D^{\otimes 4} \right\rangle \right| 1_{\{\tilde{W}^s \in A^\varepsilon \setminus A\}} ds \\ &\leq \frac{Cn}{\varepsilon^4} \int_0^1 \mathbb{E} |D|^4 1_{\{\tilde{W}^s \in A^\varepsilon \setminus A\}} ds \\ &= \frac{Cn}{\varepsilon^4} \int_0^1 \mathbb{E} |D|^4 \mathbb{P}(\tilde{W}^s \in A^\varepsilon \setminus A \mid D, U, U') ds \end{aligned}$$

Since Z is independent of D, U and U' , Lemma 4.8 yields

$$\mathbb{P}(\tilde{W}^s \in A^\varepsilon \setminus A \mid D, U, U') \leq \frac{C\varepsilon}{\tilde{\sigma}\sqrt{s}}$$

Thus we conclude

$$|R_2| \leq \frac{Cn}{\tilde{\sigma}\varepsilon^3} \mathbb{E} |D|^4 \int_0^1 \frac{1}{\sqrt{s}} ds \leq \frac{C}{\tilde{\sigma}\varepsilon^3} \sum_{i=1}^n \mathbb{E} |\xi_i|^4 \quad (64)$$

From Lemmas 4.7-4.8 and Equations (15), (62), (64), we have

$$\sup_{A \in \mathcal{B}} |\mathbb{P}(W \in A) - \mathbb{P}(Z \in A)| \leq C\tilde{\sigma}^{-1}\varepsilon + \frac{C}{\tilde{\sigma}\varepsilon} \tilde{\delta}(W, \Sigma) + \frac{C}{\tilde{\sigma}\varepsilon^3} \sum_{i=1}^n \mathbb{E} |\xi_i|^4 \quad (65)$$

and by choosing $\varepsilon = \sqrt{\tilde{\delta}(W, \Sigma)}$, we obtain

$$\sup_{A \in \mathcal{B}} |\mathbb{P}(W \in A) - \mathbb{P}(Z \in A)| \leq \frac{C}{\tilde{\sigma}} \sqrt{\tilde{\delta}(W, \Sigma)} \quad (66)$$

This completes the proof.

MATHEMATICAL MODEL OF GAN

Zhi Zhang * Xiucai Ding
Department of Statistics
UCD
wwzzhang@ucdavis.edu

1 Introduction

A recent paper ([8]) characterizes the information after a deep learning network by analyzing the eigenvalue of Gram matrix after the nonlinear neural network. They pass the Gaussian standard noise into a trained GAN model to generate three classes of images, then pass the generated data into a standard trained deep representation model CNN for multi-classification. They extracted the output, which has a large dimension, from one inner layer, analyzed their eigenvalue, and found that they have the same behavior as the GMM data constructed using the mean and variance of the GAN data. The paper considers the serials of neural network models are Lipschitz operations where the data concentrations are maintained after such operation. They conclude that given a number of the standard Gaussian vectors, which are concentrated vectors, under large n, p domain where $n, p \rightarrow \infty$ but $\frac{n}{p} \rightarrow c < \infty$, after series Lipschitz operations, the resulting matrix remain the Lipschitz concentration.

However, the following questions are unclear.

One is the relationship between the above behavior of the representation with the parameters of the GAN model. The paper generates such representation starting with the standard normal Gaussian noise vector as the concentrated vector, then passes the trained generator to obtain the representation. Due to the concentration property of standard Gaussian and Lipschitz properties of GAN operations, the resulting Gram matrix has its first and second-order statistics same as the GMM. However, such parameters of GAN may not be necessary for deducing the result.

The other unclear question is whether the distribution of noise vector matters. The standard normal Gaussian is the concentrated vector. What if we change the distribution of the noise vector? Does the result still hold?

Besides answering the above two questions, we also include an additional simulation environment, the MNIST ([6]) classification task, with a simpler model.

Our paper has the following contributions. We extend the existed paper to a new environment where the data set and neural network structures are different; we found that the findings are still valid under this new environment. In order to provide convenience

*Preprint. Work in progress. This is the project report for STA299 Independent Study In Statistics

for readers, we fully explain the mathematical model of GAN. Our paper further verifies that the eigen-behavior of Gram matrix after CNN representation by using GAN images does not matter with which distribution the noise used to generate GAN data comes from. Instead, it is an intrinsic property of the images. We also provide a comparable experiment to indicate that the values of model parameters in trained GAN do not matter; rather, the structure matters with the final Gram matrix's eigen-behavior.

2 Review of GAN

Since the birth of deep learning, it has been widely used in various fields, such as robot control, human voice authentication, biological network analysis, and image generation. However, in recent years, researchers have focused on explaining the deep learning model better.

Differentiable Generative network is one of the neural networks that transfers a latent variable z to a sample x or the distribution of x by the differentiable neural network. Generative Adversarial Network (GAN) ([3]) is belong to the differentiable generative network and is based on the game theoretical scenario where the generator is responsible for generating the training data $x^{(g)} = g(z, \theta^{(g)})$, and at the same time the discriminator is responsible for discriminating the real and generated data by maintaining a probability of the data being as real $d(x; \theta^d)$. A payoff function

$$c(\theta^g, \theta^d) = \mathbb{E}_{x^r} \log d(x^r) + \mathbb{E}_{x \sim g} \log (1 - d(x^g))$$

is defined for the discriminator so that it can better distinguish the x^r and x^g . The generator is trained upon an approach by the zero sum mean game such that

$$g^* = \underset{g}{\operatorname{argmin}} \max_d c(g, d).$$

The conditional GAN ([7]) is an extension of transitional GAN such that the overall model contains multiple GAN and each one generates data of a specific class conditionally on a class label.

Random Matrix Theory With GAN Data Their noise vector is passed through the BigGAN model ([1]) that was pre-trained from the Imagenet dataset ([2]) then followed with some popular CNN networks (Resnet50[4], VGG16[9], and Densenet201[5]). The paper GAN-data generation and representation process is a series of Lipschitz transformations such that the concentrated vectors are stable after Lipschitz transformation. The paper also proved Lipschitz's constant control over the neural network with affine layers with spectral normalization. The paper also proved that the Gram matrix and its estimation after deep learning representation is well-bounded in the large dimension domain with some assumptions.

3 Mathematical Model of GAN

Here we give the mathematical model of the GAN and CNN in our analysis.

3.1 Embedding and Conditional Random Noise Vector

Given a random noise vector $z \in \mathbb{R}^t$, and a fixed label as $l \in \mathbb{R}^{|L|}$, where t is a parameter, for example we can choose $t = 128$ by convention. l is a one hot vector to denote the class label, and L is a set of all labels. For example, we can let $|L| = 1000$. Then a given an embedding function is defined as

$$e \in \mathbb{R}^t := E(l, A) = lA^T \quad (1)$$

where $A \in \mathbb{R}^{t \times |L|}$ is a weight matrix that is initialized from $\mathcal{N}(0, 1)$. We then create a conditional vector by row-wised stacking of z and e .

$$c = (z^T, e^T) \in \mathbb{R}^d \quad (2)$$

where $d = 2t$.

The generator is a sequential functions that applied to c and we will give their formulations as below.

3.2 Linear Transformation with Spectral Normalization

A linear transformation

$$x \in \mathbb{R}^h := \mathcal{F}(c, A, b) = cA^T + b \quad (3)$$

where $A \in \mathbb{R}^{h \times d}$ is a fixed weight matrix after training, and $b \in \mathbb{R}^h$ is fixed vector of bias, which was originally randomly initialized. We can choose $h = 32678, d = 256$.

Spectral Normalization The Spectral normalization is a numerical operation applied to the A . Given the initialized raw $A_o \in \mathbb{R}^{h \times d}$ and two random vectors $u \in \mathbb{R}^h$ and $v \in \mathbb{R}^d$, the following operations are performed on A_o .

$$A = \frac{A_o}{\sigma^A} \quad (4)$$

where

$$\begin{aligned} \sigma^A &= u' \cdot (A_o v') \\ v' &= \frac{u A_o}{\max(\|u A_o\|_h, \epsilon)} \\ u' &= \frac{A_o v}{\max(\|A_o v\|_h, \epsilon)} \end{aligned}$$

Refactoring Given the previous output x after , then we refactor it to a tensor in $\mathbb{R}^{1 \times H \times W \times C}$ then permute it as $x \in \mathbb{R}^{1 \times C \times H \times W}$. Where W, H are parameters, we can choose $W = H = 4$. Then it's easy to calculate that $C = d_{11} = 2048$ in our example.

This x will be passing through series of sequential layers.

3.3 Sequential Layers

Given such x and c , the next steps are sequential layers that applied to them.

$$\mathcal{L}_n \circ \dots \circ \mathcal{L}_0(x, c, \sigma) \quad (5)$$

Where $x \in \mathbb{R}^{1 \times H \times W \times C} := \mathcal{F}(c, A, b)$. Here the \mathcal{L} could be one of any module structures as described below. Here we introduced four different modules, and the sequential layers is composed of n modules from these four categories. We will use M_1, M_2, M_3, M_4 to indicate these four modules. We also introduce another constant $\sigma \in [0, 1]$, which is also a parameter.

3.3.1 Generator Block Module Without Up Sampling and Without Drop Channels - M_1

\mathcal{L}_0 is a such module.

Batch Normalization - (M_1, \mathcal{B}_0) Here we define some pre-computed statistics for this operation. Let $f^{(M_1, B_0)}$, $s^{(M_1, B_0)}$ are the fractional and integer parts of $\frac{\sigma}{\eta}$. Let $r_{means}^{(M_1, B_0)} \in 0^{r_1^{(M_1, B_0)} \times C}$ and $r_{vars}^{(M_1, B_0)} \in 1^{r_1^{(M_1, B_0)} \times C}$, and $\eta^{(M_1, B_0)} = \frac{1}{r_1^{(M_1, B_0)} - r_2^{(M_1, B_0)}}$. And $r_{mean}^{(M_1, B_0)} = (r_{means}^{(M_1, B_0)})_{s^{(M_1, B_0), :}}$, $r_{var}^{(M_1, B_0)} = (r_{vars}^{(M_1, B_0)})_{s^{(M_1, B_0), :}}$, are sliced vectors from the matrix by setting the row index equals to $s^{(M_1, B_0)}$. Then we refactor $r_{mean}^{(M_1, B_0)} \in 0^{1 \times C \times 1 \times 1}$, $r_{var}^{(M_1, B_0)} \in 1^{1 \times C \times 1 \times 1}$.

We set hyperparameter $r_1^{(M_1, B_0)} = 51$, $r_2^{(M_1, B_0)} = 1$, and $\sigma = 1.0$, so that $s^{(M_1, B_0)} = 50$.

Then we create weight $w^{(M_1, B_0)}$ and bias $b^{(M_1, B_0)}$ by following operations: Here we let $A^{(M_1, B_0)} \in \mathbb{R}^{C \times d}$ is a spectral Normalized weight matrix. Then we apply a linear operation to c such that

$$w^{(M_1, B_0)} \in \mathbb{R}^C := 1 + cA^{(M_1, B_0)T} \quad (6)$$

and we refactor $w^{(M_1, B_0)} \in \mathbb{R}^{1 \times C \times 1 \times 1}$ and

$$b^{(M_1, B_0)} \in \mathbb{R}^C : cA^{(M_1, B_0)T} \quad (7)$$

and we refactor $b^{(M_1, B_0)} \in \mathbb{R}^{1 \times C \times 1 \times 1}$. So a batch normalization $M_1 \mathcal{B}_0$ is defined as

$$x^{(M_1, B_0)} \in \mathbb{R}^{1 \times C \times W \times H} := M_1 \mathcal{B}_0(x, r_{mean}^{(M_1, B_0)}, r_{var}^{(M_1, B_0)}, w^{(M_1, B_0)}, b^{(M_1, B_0)}) = \quad (8)$$

$$\frac{x - r_{mean}^{(M_1, B_0)}}{\sqrt{r_{var}^{(M_1, B_0)} + 0.0001}} * w^{(M_1, B_0)} + b^{(M_1, B_0)}$$

ReLU - $M_1 \mathcal{R}_0$ The rectified liner unit function is applied afterwards such that

$$x^{(M_1, R_0)} := R_0(x^{(M_1, B_0)}) = \max(0, x^{(M_1, B_0)}) \quad (9)$$

Convolution - $M_1 \mathcal{C}_0$ We let filter $A^{(M_1, C_0)}$ as the weight matrix after spectral normalization, which is a tensor in $\mathbb{R}^{C^{(M_1, C_0)} \times \frac{C}{groups} \times kH \times kW}$, and bias $b \in \mathbb{R}^{C^{(M_1, C_0)}}$, where $C^{(M_1, C_0)}$, kH , kW , $groups$ are parameters.

The convolution $M_1 \mathcal{C}_0$ is mapping from input $x^{(M_1, R_0)} \in \mathbb{R}^{1 \times C \times H \times W}$ to $x^{(M_1, C_0)} \in \mathbb{R}^{1 \times C^{(M_1, C_0)} \times H^{(M_1, C_0)} \times W^{(M_1, C_0)}}$, which is defined as:

$$x^{(M_1, C_0)} \in \mathbb{R}^{1 \times C^{(M_1, C_0)} \times H^{(M_1, C_0)} \times W^{(M_1, C_0)}} := M_1 \mathcal{C}_0(x^{(M_1, R_0)}) \quad (10)$$

where

$$H^{(M_1, C_0)} = \frac{H + 2 \times padH - dH \times (kH - 1) - 1}{sH} + 1 \quad (11)$$

$$W^{(M_1, C_0)} = \frac{W + 2 \times padW - dW \times (kW - 1) - 1}{sW} + 1 \quad (12)$$

For example, we have $C^{(M_1, L_0, C_0)} = d_4 = 512$, $C_{in} = d_{16} = 2048$, $groups = 1$, we also add following parameters: let $iH = 4$, $iW = 4$, $kH = 1$, $kW = 1$, $sH = 1$, $sW = 1$, $dH = 1$, $dW = 1$, it means $stride = 1$, and $padding = 0$, and $dilation = 1$. Then the $H^{(M_1, L_0, C_0)} = W^{(M_1, L_0, C_0)} = 4$

Batch Normalization M_1, \mathcal{B}_1 and Relu M_1, R_1 It has the same operations as $M_1 \mathcal{B}_0$, so that

$$x^{(M_1, B_1)} \in \mathbb{R}^{1 \times C^{(M_1, C_0)} \times H^{(M_1, C_0)} \times W^{(M_1, C_0)}} := \mathcal{B}_1(x^{(M_1, C_0)}, c, \sigma) \quad (13)$$

$$x^{(M_1, R_1)} \in \mathbb{R}^{1 \times C^{(M_1, C_0)} \times H^{(M_1, C_0)} \times W^{(M_1, C_0)}} := R_1(x^{(M_1, B_1)}) = \max(0, x^{M_1, B_1}) \quad (14)$$

where $C^{(M_1, C_0)} = d_4 = 512$

Convolution M_1, \mathcal{C}_1 We let filter $A^{(M_1, C_1)}$ as the weight matrix after spectral normalization, which is a tensor in $\mathbb{R}^{C^{(M_1, C_1)} \times \frac{C^{(M_1, C_0)}}{groups} \times kH \times kW}$, and let bias as $b^{(M_1, C_1)} \in \mathbb{R}^{C^{(M_1, C_1)}}$,

The $M_1 \mathcal{C}_1$ is a mapping from input $x^{(M_1 R_1)} \in \mathbb{R}^{1 \times C^{(M_1 C_0)} \times H^{(M_1 C_0)} \times W^{(M_1 C_0)}}$ to $x^{(M_1 C_1)} \in \mathbb{R}^{1 \times C^{(M_1 C_1)} \times H^{(M_1 C_1)} \times W^{(M_1 C_1)}}$ defined as:

$$x^{(M_1, C_1)} \in \mathbb{R}^{1 \times C^{(M_1, C_1)} \times H^{(M_1, C_1)} \times W^{(M_1, C_1)}} := M_1 \mathcal{C}_1(x^{(M_1, R_1)}) \quad (15)$$

Here we let $A^{(M_1 C_1)} \in \mathbb{R}^{512 \times 512 \times 3 \times 3}$ as, we let $b^{(M_1 C_1)} \in \mathbb{R}^{512}$. So $C_{out} = d_4 = 512$, $C_{in} = d_4 = 512$. For our calculation, we also add following variables, $iH = 4$, $iW = 4$, $kH = 1$, $kW = 1$, $sH = 1$, $sW = 1$, $dH = 1$, $dW = 1$, $padH = 1$, $padW = 1$, it means $kernelsize = 3$, $stride = 1$, and $padding = 1$, and $dilation = 1$, and $groups = 1$. Then the $H^{(C_1)} = W^{(C_1)} = 4$

Batch Normalization $M_1 \mathcal{B}_2$ and ReLU $M_1 \mathcal{R}_2$ and Convolution $M_1 \mathcal{C}_2$ It has the same operations as $M_1 \mathcal{B}_0$, $M_1 \mathcal{B}_1$, so that

$$x^{(M_1 B_2)} \in \mathbb{R}^{1 \times C^{(M_1 C_1)} \times H^{(M_1 C_1)} \times W^{(M_1 C_1)}} := M_1 \mathcal{B}_2(x^{(M_1 C_1)}, c, \sigma) \quad (16)$$

$$x^{(M_1 R_2)} := M_1 \mathcal{R}_2(x^{(M_1 B_2)}) \max(0, x^{(M_1 B_2)}) \quad (17)$$

$$x^{(M_1 C_2)} \in \mathbb{R}^{1 \times C^{(M_1 C_2)} \times H^{(M_1 C_2)} \times W^{(M_1 C_2)}} := M_1 \mathcal{C}_2(x^{(M_1 R_2)}) \quad (18)$$

where $C^{(M_1 C_2)} = 512$, $H^{(M_1 C_2)} = W^{(M_1 C_2)} = 4$

Batch Normalization $M_1 \mathcal{B}_3$ and ReLU $M_1 \mathcal{R}_3$ It has the same operations as $M_1 \mathcal{B}_0$, $M_1 \mathcal{B}_1$, $M_1 \mathcal{B}_2$, so that

$$x^{(M_1, B_3)} \in \mathbb{R}^{1 \times C^{(M_1, C_2)} \times H^{(M_1, C_2)} \times W^{(M_1, C_2)}} := \mathcal{B}_3(x^{(M_1, C_2)}, c, \sigma) \quad (19)$$

$$x^{(M_1, R_3)} := M_1 \mathcal{R}_3(x^{(M_1, B_3)}) = \max(0, x^{(M_1, B_3)}) \quad (20)$$

Convolution - M_1, \mathcal{C}_3 Let filter $A^{(M_1, C_3)}$ as the weight matrix after spectral normalization, a tensor in $\mathbb{R}^{C^{M_1, C_3} \times \frac{C^{M_1, C_2}}{groups} \times kH \times kW}$, And bias $b^{(M_1, C_3)} \in \mathbb{R}^{C^{M_1, C_3}}$,

The convolution - $M_1 \mathcal{C}_3$ is mapping from input $x^{(M_1, C_2)} \in \mathbb{R}^{1 \times C^{(M_1, C_2)} \times H^{(M_1, C_2)} \times W^{(M_1, C_2)}}$ to $x^{(M_1, C_3)} \in \mathbb{R}^{1 \times C^{(M_1, C_3)} \times H^{(M_1, C_3)} \times W^{(M_1, C_3)}}$ defined as:

$$x^{(M_1, C_3)} \in \mathbb{R}^{1 \times C^{M_1, C_3} \times H^{(M_1, C_3)} \times W^{(M_1, C_3)}} := M_1 \mathcal{C}_3(x^{(M_1, R_3)}) \quad (21)$$

here we let $A^{(M_1, R_3)} \in \mathbb{R}^{2048 \times 512 \times 1 \times 1}$, we let $b^{(M_1, R_3)} \in \mathbb{R}^{2048}$. Here we let $C^{(M_1, C_3)} = d_{11} = 2048$, and $C^{(M_1, C_2)} = d_9 = 512$. For our calculation, we also add following variable $H = 1, W = 1, H^{(C_2)} = 4, W^{(C_2)} = 4, kH = 1, kW = 1, sH = 1, sW = 1, dH = 1, dW = 1, padH = 0, padW = 0$, it means $kernelsize = 1, stride = 1$, and $padding = 1$, and $dilation = 1$, and $groups = 1$. And $H^{(C_3)} = W^{(C_3)} = 4$

Addition - $M_1\mathcal{A}_0$ The addition operation is

$$x^{(M_1, A_0)} \in \mathbb{R}^{1 \times C^{(M_1, C_3)} \times H^{(M_1, C_3)} \times W^{(M_1, C_3)}} := x^{(M_1, C_3)} + x \quad (22)$$

We have the other three types of modules defined in the Appendix ??

3.3.2 Summary

For notation wise, we have used M_1, M_2, M_3, M_4 to denote the above four different modules. And the next module's input channel size is the previous module's output channel size.

So the x and c pass through series modules, we give the details in the following:

$$x^{M_1, L_0} \in \mathbb{R}^{1 \times C^{(M_1, L_0, C_3)} \times H^{(M_1, L_0, C_3)} \times W^{(M_1, L_0, C_3)}} := M_1\mathcal{L}_0(x \in \mathbb{R}^{1 \times C^C \times H^C \times W^C}, c \in \mathbb{R}^{1 \times d}, \sigma) \quad (23)$$

where $C^C = 2048, H^C = W^C = 4, C^{(M_1, L_0, C_3)} = 2048, H^{(M_1, L_0, C_3)} = W^{(M_1, L_0, C_3)} = 4$

$$x^{M_2, L_1} \in \mathbb{R}^{1 \times C^{(M_2, L_1, C_3)} \times H^{(M_2, L_1, C_3)} \times W^{(M_2, L_1, C_3)}} := M_2\mathcal{L}_1(x, c, \sigma) \quad (24)$$

$$C^{(M_2, L_1, C_3)} = 2048, H^{(M_2, L_1, C_3)} = W^{(M_2, L_1, C_3)} = 8$$

$$x^{M_1, L_2} \in \mathbb{R}^{1 \times C^{(M_1, L_2, C_3)} \times H^{(M_1, L_2, C_3)} \times W^{(M_1, L_2, C_3)}} := M_1\mathcal{L}_2(x, c, \sigma) \quad (25)$$

$$C^{(M_1, L_2, C_3)} = 2048, H^{(M_1, L_2, C_3)} = W^{(M_1, L_2, C_3)} = 8$$

$$x^{M_3, L_3} \in \mathbb{R}^{1 \times C^{(M_3, L_3, C_3)} \times H^{(M_3, L_3, C_3)} \times W^{(M_3, L_3, C_3)}} := M_3\mathcal{L}_3(x, c, \sigma) \quad (26)$$

$$C^{(M_3, L_3, C_3)} = 1024, H^{(M_3, L_3, C_3)} = W^{(M_3, L_3, C_3)} = 16$$

$$x^{M_1, L_4} \in \mathbb{R}^{1 \times C^{(M_1, L_4, C_3)} \times H^{(M_1, L_4, C_3)} \times W^{(M_1, L_4, C_3)}} := M_1\mathcal{L}_4(x, c, \sigma) : \quad (27)$$

$$C^{(M_1, L_4, C_3)} = 1024, H^{(M_1, L_4, C_3)} = W^{(M_1, L_4, C_3)} = 16$$

$$x^{M_2, L_5} \in \mathbb{R}^{1 \times C^{(M_2, L_5, C_3)} \times H^{(M_2, L_5, C_3)} \times W^{(M_2, L_5, C_3)}} := M_2\mathcal{L}_5(x, c, \sigma) \quad (28)$$

$$C^{(M_2, L_5, C_3)} = 1024, H^{(M_2, L_5, C_3)} = W^{(M_2, L_5, C_3)} = 32$$

$$x^{M_1, L_6} \in \mathbb{R}^{1 \times C^{(M_1, L_6, C_3)} \times H^{(M_1, L_6, C_3)} \times W^{(M_1, L_6, C_3)}} := M_1 \mathcal{L}_6(x, c, \sigma) \quad (29)$$

$$C^{(M_1, L_6, C_3)} = 1024, H^{(M_1, L_6, C_3)} = W^{(M_1, L_6, C_3)} = 32$$

$$x^{M_3, L_7} \in \mathbb{R}^{1 \times C^{(M_3, L_7, C_3)} \times H^{(M_3, L_7, C_3)} \times W^{(M_3, L_7, C_3)}} := M_3 \mathcal{L}_7(x, c, \sigma) \quad (30)$$

$$C^{(M_3, L_7, C_3)} = 512, H^{(M_3, L_7, C_3)} = W^{(M_3, L_7, C_3)} = 64$$

$$x^{M_4, L_8} \in \mathbb{R}^{1 \times C^{(M_4, L_8, C_3)} \times H^{(M_4, L_8, C_3)} \times W^{(M_4, L_8, C_3)}} := M_4 \mathcal{L}_8(x) \quad (31)$$

$$C^{(M_4, L_8, C_3)} = 512, H^{(M_4, L_8, C_3)} = W^{(M_4, L_8, C_3)} = 64$$

$$x^{M_1, L_9} \in \mathbb{R}^{1 \times C^{(M_1, L_9, C_3)} \times H^{(M_1, L_9, C_3)} \times W^{(M_1, L_9, C_3)}} := M_1 \mathcal{L}_9(x, c, \sigma) \quad (32)$$

$$C^{(M_1, L_9, C_3)} = 512, H^{(M_1, L_9, C_3)} = W^{(M_1, L_9, C_3)} = 64$$

$$x^{M_3, L_{10}} \in \mathbb{R}^{1 \times C^{(M_3, L_{10}, C_3)} \times H^{(M_3, L_{10}, C_3)} \times W^{(M_3, L_{10}, C_3)}} := M_3 \mathcal{L}_{10}(x, c, \sigma) \quad (33)$$

$$C^{(M_3, L_{10}, C_3)} = 256, H^{(M_3, L_{10}, C_3)} = W^{(M_3, L_{10}, C_3)} = 128$$

$$x^{M_1, L_{11}} \in \mathbb{R}^{1 \times C^{(M_1, L_{11}, C_3)} \times H^{(M_1, L_{11}, C_3)} \times W^{(M_1, L_{11}, C_3)}} := M_1 \mathcal{L}_{11}(x, c, \sigma) \quad (34)$$

$$C^{(M_1, L_{11}, C_3)} = 256, H^{(M_1, L_{11}, C_3)} = W^{(M_1, L_{11}, C_3)} = 128$$

$$x^{M_3, L_{12}} \in \mathbb{R}^{1 \times C^{(M_3, L_{12}, C_3)} \times H^{(M_3, L_{12}, C_3)} \times W^{(M_3, L_{12}, C_3)}} := M_3 \mathcal{L}_{12}(x, c, \sigma) \quad (35)$$

$$C^{(M_3, L_{12}, C_3)} = 128, H^{(M_3, L_{12}, C_3)} = W^{(M_3, L_{12}, C_3)} = 256$$

So in the end, we have sequential layers operator \mathcal{L} defined as

$$\mathcal{L}(x, c, \sigma) : M_3 \mathcal{L}_{12} \circ \dots \circ M_1 \mathcal{L}_0(x, c, \sigma) \quad (36)$$

3.4 Batch Normalization B with ReLU R

The last batch normalization is applied to the output after sequential layers. We let the input of B as $x^{iB} \in \mathbb{R}^{1 \times C^{cB} \times H^{cB} \times W^{cB}}$. Here we define some pre-computed statistics for this operation. And the input channel size is $C^{(L_{12}, C_3)}$. Let $f^{(B)}, s^{(B)}$ are the fractional and integer parts of $\frac{\sigma}{\eta}$. Let $r_{means}^{(B)} \in 0^{r_1^{(B)} \times C^{cB}}$ and $r_{vars}^{(B)} \in 1^{r_1^{(B)} \times C^{cB}}$, and $\eta = \frac{1}{r_1^{(B)} - r_2^{(B)}}$. And $r_{mean}^{(B)} = (r_{means}^{(B)})_{s(B), :}$, $r_{var}^{(B)} = (r_{vars}^{(B)})_{s(B), :}$ are sliced

vectors from the matrix by setting the row index equals to $s^{(B)}$. Then we refactor $r_{mean}^{(B)} \in \mathbb{R}^{1 \times C^{cB} \times 1 \times 1}$, $r_{var}^{(B)} \in \mathbb{R}^{1 \times C^{cB} \times 1 \times 1}$.

Unlike the \mathcal{B} which is a conditional batch normalization considering the c , The \mathcal{B} create weight w^B and bias b^B by random initialization.

$$w^B \in \mathbb{R}^{1 \times C^{cB}} \quad (37)$$

and we refactor $w^B \in \mathbb{R}^{1 \times C^{cB} \times 1 \times 1}$ and

$$b^B \in \mathbb{R}^{1 \times C^{cB}} \quad (38)$$

and we refactor $b^B \in \mathbb{R}^{1 \times C^{cB} \times 1 \times 1}$.

And after training, we have the fixed value for w^B, b^B . So a batch normalization \mathcal{B} is defined as

$$x^{(B)} \in \mathbb{R}^{1 \times C^{cB} \times H^{(cB)} \times W^{(cB)}} := \mathcal{B}(x^{iB}) = \frac{x^{iB} - r_{mean}}{\sqrt{r_{var} + \epsilon}} * w^B + b^B \quad (39)$$

ReLU - \mathcal{R} The rectified liner unit function is applied afterwards such that

$$x^{(R)} \in \mathbb{R}^{1 \times 128 \times 256 \times 256} := \mathcal{R}(x^{(B)}) : \max(0, x^{(B)}) \quad (40)$$

For our example, our $iB = (M_3, L_{12}, A_0)$, $cB = (M_3, L_{12}, C_3)$. We choose $\epsilon = 0.0001$. We set hyperparameter $r_1^{(B)} = 51, r_2^{(B)} = 1$, where $C^{(M_3, L_{12}, C_3)} = 128, H^{M_3, L_{12}, C_3} = W^{(M_3, L_{12}, C_3)} = 256$

3.5 Convolution with RGB

Convolution - \mathcal{C}_o Let filter $A^{(C_o)}$ as the weight matrix after spectral normalization, which is a tensor in $\mathbb{R}^{C^{C_o} \times \frac{C^{(iB_c)}}{groups} \times kH \times kW}$. And bias $b \in \mathbb{R}^{C^{C_o}}$, And the input $x^{(R)} \in \mathbb{R}^{1 \times C^{(cB)} \times H^{(cB)} \times W^{(cB)}}$.

The convolution with RGB - \mathcal{C}_o is a mapping from input $x^{(R)}$ to $x^{(C_o)} \in \mathbb{R}^{1 \times C^{(C_o)} \times H^{(C_o)} \times W^{(C_o)}}$ defined as:

$$x^{(C_o)} \in \mathbb{R}^{1 \times C^{(C_o)} \times H^{(C_o)} \times W^{(C_o)}} := \mathcal{C}_o(x^{(R)}) \quad (41)$$

Here we let $A^{(C_o)} \in \mathbb{R}^{128 \times 128 \times 3 \times 3}$ we let $b^{(C_o)} \in \mathbb{R}^{128}$. height and weight can be: $iH = 64, iW = 64, C_{in} = 128$, the kernel height and weight can be $kH = 1, kW = 1, sH = 1, sW = 1, dH = 1, dW = 1, padH = 1, padW = 1$, it means $kernelsize = 1, stride = 1$, and $padding = 1$, and $dilation = 1$, and $groups = 1$. We have $C^{(C_o)} = 128, H^{(C_o)} = W^{(C_o)} = 256$.

Slicing For the $C^{(C_o)}$ channels in $x^{(C_o)}$, we select the first 3 channel from it to represent RGB. The slicing operator $\mathcal{S}l$ is:

$$x^{(Sl)} \in \mathbb{R}^{1 \times C^{(Sl)} \times H^{(Sl)} \times W^{(Sl)}} := \mathcal{S}(x^{(C_o)}) \quad (42)$$

where $C^{(Sl)} = 3$, and $H^{(Sl)} = W^{(Sl)} = 256$

3.6 Tanh activation

The last operation is the hyperbolic tangent tanh activation \tanh such that

$$x^{(Th)} \in \mathbb{R}^{1 \times C^{(Sl)} \times H^{(Sl)} \times W^{(Sl)}} := \text{Tanh}(x) = \frac{\exp(x) - \exp(-x)}{\exp(x) + \exp(-x)} \quad (43)$$

So the output value is between $[-1, 1]$

4 Convert to Images

This section we convert the $x^{(Th)}$ into actual images.

4.1 Refactor

The refactor Ref_o is a operation that reorders x such that

$$x^{(Ref)} \in \mathbb{R}^{1 \times H^{(Sl)} \times W^{(Sl)} \times C^{(Sl)}} := Ref(x^{Th}) \quad (44)$$

Then the original value is converted to $[0, 255]$ by

$$x_{gan} = (x^{(Ref)} + 1)/2 \times 256$$

. And we control the range of x_{gan} falls into $[0, 255]$

5 Mathematical Model of Convolution layers Resnet-50

5.1 Transformer

The transformer re-scale, and normalize the original image. The original image is in RGB format that has dimension $\mathbb{R}^{H^{(Sl)} \times W^{(Sl)} \times C^{(Sl)}}$.

Resize - Rs Resize transformer Rs is defined such that:

$$x^{(Rs)} \in \mathbb{R}^{C^{(Sl)} \times H^{(Rs)} \times W^{(Rs)}} := Rs(x_{gan}) \quad (45)$$

where $H^{(Rs)} = W^{(Rs)} = 224, C^{(Sl)} = 3$

Normalization A normalization \mathcal{N} is a operator that applies to $x^{(Rs)}$ in each channel such that:

$$x^{(N)} := \mathcal{N}(x^{(Rs)}) = \frac{x^{(Rs)} - \text{mean}(x^{(Rs)})}{\text{std}(x^{(Rs)})} \quad (46)$$

where $\text{mean}(x^{(Rs)}), \text{std}(x^{(Rs)}) \in \mathbb{R}^{C^{(Sl)} \times 1 \times 1}$ Then we refactor $x^{(N)} \in \mathbb{R}^{1 \times C^{(Sl)} \times H^{(Rs)} \times W^{(Rs)}}$

A embedding $e \in 0^{1 \times C \times 1 \times 1}$ will be used. where $C = 2048$

5.2 Convolution, batch normalization, relu, max-pooling layers

Convolution - \mathcal{C}_0^{Res-50} Let filter A as the weight matrix, which is a tensor in $\mathbb{R}^{C^{(C_0^{Res-50})} \times \frac{C^{Sl}}{groups}, kH, kW}$, And the input is $x^{(N)} \in \mathbb{R}^{batch \times C^{(Sl)} \times H^{(Rs)} \times W^{(Rs)}}$,

The convolution \mathcal{C}_0^{Res-50} is mapping from input $x^{(N)} \in \mathbb{R}^{1 \times C^{(Sl)} \times H^{(Rs)} \times W^{(Rs)}}$ to $x \in \mathbb{R}^{N \times C^{(C_0^{Res-50})} \times H^{(C_0^{Res-50})} \times W^{(C_0^{Res-50})}}$ defined as:

$$x^{(C_0^{Res-50})} \in \mathbb{R}^{1 \times 64 \times 112 \times 112} := \mathcal{C}_0^{Res-50}(x) \quad (47)$$

here we let $A^{(C_0^{Res-50})} \in \mathbb{R}^{64 \times 3 \times 7 \times 7}$. height and weight can be: $H^{(Rs)} = 254, W^{(Rs)} = 254, C^{(Sl)} = 3$, the kernel height and weight can be $kH = 7, kW = 7, sH = 2, sW = 2, dH = 1, dW = 1, padH = 3, padW = 3$, it means $kernelsize = 7$, $stride = 2$, and $padding = 3$, and $dilation = 1$, and $groups = 1$, and $C^{C_0^{Res-50}} = 64, H^{C_0^{Res-50}} = W^{C_0^{Res-50}} = 112$

Batch Normalization - \mathcal{B}_0^{Res-50} The last batch normalization is applied to the $x^{(C_0^{Res-50})}$. The \mathcal{B}_0^{Res-50} use pre-computed statistics. $r_{means}^{\mathcal{B}_0^{Res-50}} \in 0^{C_0^{Res-50}}$ and $r_{vars}^{\mathcal{B}_0^{Res-50}} \in 1^{C_0^{Res-50}}$, and $momentum = 0.1$. After training, $r_{means}^{\mathcal{B}_0^{Res-50}}, r_{vars}^{\mathcal{B}_0^{Res-50}}$ are returned from training data. the So $r_{mean}^{\mathcal{B}_0^{Res-50}} = momentum * input_{means} + (1 - momentum) * r_{means}^{\mathcal{B}_0^{Res-50}}$, $r_{var}^{\mathcal{B}_0^{Res-50}} = momentum * input_{vars} + (1 - momentum) * r_{vars}^{\mathcal{B}_0^{Res-50}}$. Then we refactor $r_{mean}^{\mathcal{B}_0^{Res-50}} \in 0^{1 \times C_0^{Res-50} \times 1 \times 1}, r_{var}^{\mathcal{B}_0^{Res-50}} \in 1^{1 \times C_0^{Res-50} \times 1 \times 1}$.

The input channel size is 64. And $\epsilon = 0.00001$

The \mathcal{B}_0^{Res-50} create weight $w^{\mathcal{B}_0^{Res-50}}$ and bias $b^{\mathcal{B}_0^{Res-50}}$ by random initialization.

$$w^{\mathcal{B}_0^{Res-50}} \in 0^{1 \times C^{(C_0^{Res-50})}} \quad (48)$$

and we refactor $w^{\mathcal{B}_0^{Res-50}} \in \mathbb{R}^{1 \times C^{(C_0^{Res-50})} \times 1 \times 1}$ and

$$b^{\mathcal{B}_0^{Res-50}} \in 0^{1 \times C^{(C_0^{Res-50})}} \quad (49)$$

and we refactor $b^{\mathcal{B}_0^{Res-50}} \in \mathbb{R}^{1 \times C^{(C_0^{Res-50})} \times 1 \times 1}$.

And after training, we have the fixed value for $w^{\mathcal{B}_0^{Res-50}}, b^{\mathcal{B}_0^{Res-50}}$. So a batch normalization \mathcal{B}_0^{Res-50} is defined as

$$x^{(\mathcal{B}_0^{Res-50})} \in \mathbb{R}^{1 \times C^{(C_0^{Res-50})} \times H^{(C_0^{Res-50})} \times W^{(C_0^{Res-50})}} := \mathcal{B}_0^{Res-50}(x^{(C_0^{Res-50})}) = \frac{x^{(C_0^{Res-50})} - r_{mean}^{(C_0^{Res-50})}}{\sqrt{r_{var}^{(C_0^{Res-50})}} + \epsilon} * w^{(C_0^{Res-50})} + b^{(C_0^{Res-50})} \quad (50)$$

Relu - R_0^{Res-50} The rectified liner unit function is applied afterwards such that

$$x^{R_0^{Res-50}} \in \mathbb{R}^{1 \times C^{(C_0^{Res-50})} \times H^{(C_0^{Res-50})} \times W^{(C_0^{Res-50})}} := R_0^{Res-50}(x^{(\mathcal{B}_0^{Res-50})}) = max(0, x^{(\mathcal{B}_0^{Res-50})}) \quad (51)$$

Max Pooling Layer - MP_0 A max pooling layer is followed after the R_0^{Res-50} . Here we have input $x^{R_0^{Res-50}} \in \mathbb{R}^{1 \times C^{(C_0^{Res-50})} \times H^{(C_0^{Res-50})} \times W^{(C_0^{Res-50})}}$. A max pooling layer is function mapping from input $x^{(R_0^{Res-50})} \in \mathbb{R}^{1 \times C^{(C_0^{Res-50})} \times H^{(C_0^{Res-50})} \times W^{(C_0^{Res-50})}}$ to $x^{(MP_0^{Res-50})} \in \mathbb{R}^{1 \times C^{(MP_0^{Res-50})} \times H^{(MP_0^{Res-50})} \times W^{(MP_0^{Res-50})}}$ defined as:

$$x^{(MP_0^{Res-50})} \in \mathbb{R}^{1 \times C^{(MP_0^{Res-50})} \times H^{(MP_0^{Res-50})} \times W^{(MP_0^{Res-50})}} := MP_0(x^{(R_0^{Res-50})}) \quad (52)$$

where

$$H^{(MP_0^{Res-50})} = \frac{H^{(C_0^{Res-50})} + 2 \times padH - dH \times (kH - 1) - 1}{sH} + 1 \quad (53)$$

$$W^{(MP_0^{Res-50})} = \frac{W^{(C_0^{Res-50})} + 2 \times padW - dW \times (kW - 1) - 1}{sW} + 1 \quad (54)$$

We also have the kernel height and weight be $kH = 3, kW = 3, sH = 2, sW = 2, dH = 1, dW = 1, padH = 1, padW = 1$, it means *kernelseize* = 3, *stride* = 2, and *padding* = 1, and *dilation* = 1, and *groups* = 1, where $C^{(MP_0^{Res-50})} = 64$, and $H^{(MP_0^{Res-50})} = W^{(MP_0^{Res-50})} = 56$

5.3 Sequential Layers

One sequential layers \mathcal{S} could contain the following layers:

A bottleneck with down-sample layers, a convolution layer and a batch normalization layer and a down-sample layer, then a relu layer. For simplicity, we give this kind of bottleneck as \mathcal{L}^d .

A bottleneck without down-sample layers, a convolution layer and a batch normalization layer, then a relu layer. We give this kind of bottleneck as \mathcal{L}^b .

Convolution layer - C_1^{Res-50} Let filter $A^{(C_1^{Res-50})}$ as the weight matrix, which is a tensor in $\mathbb{R}^{C^{(C_1^{Res-50})} \times \frac{cC_1^{Res-50}}{groups}, kH, kW}$, And the input could be the output of max pooling layer, or the previous sequential layer. So we use $x^{(iC_1^{Res-50})} \in \mathbb{R}^{1 \times C^{(cC_1^{Res-50})} \times H^{(cC_1^{Res-50})} \times W^{(cC_1^{Res-50})}}$ to implicitly represent the input.

The convolution \mathcal{C}_1^{Res-50} is mapping from input $x^{(iC_1^{Res-50})}$ to $x \in \mathbb{R}^{1 \times C^{(C_1^{Res-50})} \times H^{(C_1^{Res-50})} \times W^{(C_1^{Res-50})}}$ defined as:

$$x^{(C_1^{Res-50})} \in \mathbb{R}^{1 \times C^{(C_1^{Res-50})} \times H^{(C_1^{Res-50})} \times W^{(C_1^{Res-50})}} := \mathcal{C}_1^{Res-50}(x^{(C_1^{Res-50})}) \quad (55)$$

Suppose the input is the $x^{(MP_0^{Res-50})}$ here we let $A^{(C_1^{Res-50})} \in \mathbb{R}^{64 \times 64 \times 1 \times 1}$, we let height and weight can be: $H^{(cC_1^{Res-50})} = 56, W^{(cC_1^{Res-50})} = 56, C^{(cC_1^{Res-50})} = 64$, the kernel height and weight can be $kH = 1, kW = 1, sH = 1, sW = 1, dH = 1, dW = 1, padH = 0, padW = 0$, it means *kernelseize* = 1, *stride* = 1, and *padding* = 0, and *dilation* = 1, and *groups* = 1. We have $C^{(C_1^{Res-50})} = 64$ and $H^{(C_1^{Res-50})} = W^{(C_1^{Res-50})} = 56$

Batch Normalization layer - \mathcal{B}_1^{Res-50} The last batch normalization is applied to the $x^{(C_1^{Res-50})}$. The \mathcal{B}_1^{Res-50} use pre-computed statistics. The input channel size is $C^{(C_1^{Res-50})}$. $r_{means} \in 0^{C^{(C_1^{Res-50})}}$ and $r_{vars} \in 1^{C^{(C_1^{Res-50})}}$, and *momentum* = 0.1. After training, r_{means}, r_{vars} are returned from training data. the So $r_{mean} = momentum * input_{means} + (1 - momentum) * r_{means}$, $r_{var} = momentum * input_{vars} + (1 - momentum) * r_{vars}$. Then we refactor $r_{mean} \in 0^{1 \times (C_1^{Res-50}) \times 1 \times 1}$, $r_{var} \in 1^{1 \times C^{(C_1^{Res-50})} \times 1 \times 1}$. And $\epsilon = 0.00001$

The \mathcal{B}_1^{Res-50} create weight $w^{\mathcal{B}_1^{Res-50}}$ and bias $b^{\mathcal{B}_1^{Res-50}}$ by random initialization.

$$w^{\mathcal{B}_1^{Res-50}} \in \mathbb{R}^{1 \times C_1^{Res-50}} \quad (56)$$

and we refactor $w^{\mathcal{B}_1^{Res-50}} \in \mathbb{R}^{1 \times C_1^{Res-50} \times 1 \times 1}$ and

$$b^{\mathcal{B}_1^{Res-50}} \in \mathbb{R}^{1 \times C_1^{Res-50}} \quad (57)$$

and we refactor $b^{\mathcal{B}_1^{Res-50}} \in \mathbb{R}^{1 \times C_1^{Res-50} \times 1 \times 1}$.

And after training, we have the fixed value for $w^{\mathcal{B}_1^{Res-50}}, b^{\mathcal{B}_1^{Res-50}}$. So a batch normalization \mathcal{B}_1^{Res-50} is defined as

$$x^{(B_1^{Res-50})} \in \mathbb{R}^{1 \times C_1^{Res-50} \times H^{C_1^{Res-50}} \times W^{C_1^{Res-50}}} := \mathcal{B}(x^{(C_1^{Res-50})}) = \frac{x^{(C_1^{Res-50})} - r_{mean}}{\sqrt{r_{var} + \epsilon}} * w^{\mathcal{B}_1^{Res-50}} + b^{\mathcal{B}_1^{Res-50}} \quad (58)$$

ReLU layer - R_1^{Res-50} The rectified liner unit function is applied afterwards such that

$$x^{R_1^{Res-50}} \in \mathbb{R}^{1 \times 64 \times 56 \times 56} := R_1^{Res-50}(x^{(B_1^{Res-50})}) = \max(0, x^{(B_1^{Res-50})}) \quad (59)$$

Convolution layer - $\mathcal{C}_2^{(Res-50)}$ Let filter $A^{\mathcal{C}_2^{(Res-50)}}$ as the weight matrix, which is a tensor in $\mathbb{R}^{C_2^{(Res-50)} \times \frac{C_1^{(Res-50)}}{groups} \times kH \times kW}$, And the input $x^{R_1^{Res-50}} \in \mathbb{R}^{1 \times C_1^{(Res-50)} \times H^{C_1^{(Res-50)}} \times W^{C_1^{(Res-50)}}}$.

The convolution \mathcal{C} is mapping from input $x^{R_1^{Res-50}} \in \mathbb{R}^{1 \times C_{in} \times H^{C_1^{Res-50}} \times W^{C_1^{Res-50}}}$ to $x \in \mathbb{R}^{1 \times C_2^{(Res-50)} \times H^{C_2^{(Res-50)}} \times W^{C_2^{(Res-50)}}}$ defined as:

$$x^{(\mathcal{C}_2^{(Res-50)})} \in \mathbb{R}^{1 \times 64 \times 56 \times 56} := \mathcal{C}(x^{R_1^{Res-50}}) \quad (60)$$

Here we let $A^{\mathcal{C}_2^{(Res-50)}} \in \mathbb{R}^{64 \times 64 \times 3 \times 3}$. The input height and weight can be: $iH = 56, iW = 56, C_{in} = 64$, the kernel height and weight can be $kH = 3, kW = 3, sH = 1, sW = 1, dH = 1, dW = 1, padH = 1, padW = 1$, it means $kernelsize = 3, stride = 1$, and $padding = 1$, and $dilation = 1$, and $groups = 1$. We have $C_2^{(Res-50)} = 64$, and $H^{C_2^{(Res-50)}} = W^{C_2^{(Res-50)}} = 56$

Batch Normalization layer - \mathcal{B}_2^{Res-50} Same operation as previous batch normalization layer, here a batch normalization \mathcal{B}_2^{Res-50} is defined as

$$x^{(\mathcal{B}_2^{Res-50})} \in \mathbb{R}^{1 \times 64 \times 56 \times 56} := \mathcal{B}(x^{(\mathcal{C}_2^{Res-50})}) = \frac{x^{(\mathcal{C}_2^{Res-50})} - r_{mean}}{\sqrt{r_{var} + \epsilon}} * w^{(\mathcal{B}_2^{Res-50})} + b^{(\mathcal{B}_2^{Res-50})} \quad (61)$$

Relu layer - $\mathcal{R}_2^{(Res-50)}$ The rectified liner unit function is applied afterwards such that

$$x^{(R_2^{(Res-50)})} \in \mathbb{R}^{1 \times C_2^{(Res-50)} \times H^{C_2^{(Res-50)}} \times W^{C_2^{(Res-50)}}} := \mathcal{R}_2^{(Res-50)}(x^{(\mathcal{B}_2^{Res-50})}) : \max(0, x^{(\mathcal{B}_2^{Res-50})}) \quad (62)$$

Convolution layer - \mathcal{C}_3^{Res-50} Let filter $A^{(C_3^{Res-50})}$ be the weight matrix, which a tensor in $\mathbb{R}^{C(C_3^{Ret-50}) \times \frac{C(C_2^{Res-50})}{groups} \times kH \times kW}$,

The convolution \mathcal{C}_3^{Res-50} is mapping from input $x^{(R_2^{(Res-50)})} \in \mathbb{R}^{1 \times C(C_3^{Ret-50}) \times H(C_3^{Res-50}) \times W(C_3^{Res-50})}$ to $x \in \mathbb{R}^{1 \times C(C_3^{Ret-50}) \times H(C_3^{Res-50}) \times W(C_3^{Res-50})}$ defined as:

$$x^{(C_3^{Ret-50})} \in \mathbb{R}^{1 \times C(C_3^{Ret-50}) \times H(C_3^{Ret-50}) \times W(C_3^{Ret-50})} := \mathcal{C}(x^{(R_2^{(Res-50)})}) \quad (63)$$

here we let $A^{(C_3^{Ret-50})} \in \mathbb{R}^{256 \times 64 \times 1 \times 1}$. And the input is in $\mathbb{R}^{1 \times C(C_2^{Res-50}) \times H(C_2^{Res-50}) \times W(C_2^{Res-50})}$ height and weight can be: $iH = 56, iW = 56, C_{in} = 64$, the kernel height and weight can be $kH = 1, kW = 1, sH = 1, sW = 1, dH = 1, dW = 1, padH = 0, padW = 0$, it means $kernelsize = 1, stride = 1$, and $padding = 1$, and $dilation = 1$, and $groups = 1$. We have $C(C_3^{Ret-50}) = 256, H(C_3^{Ret-50}) = W(C_3^{Ret-50}) = 56$

Batch Normalization layer - \mathcal{B}_3^{Res-50} Same operation as previous batch normalization layer, here a batch normalization \mathcal{B}_3^{Res-50} is defined as

$$x^{(B_3^{Res-50})} \in \mathbb{R}^{1 \times 256 \times 56 \times 56} := \mathcal{B}_3^{Res-50} : \frac{x - r_{mean}}{\sqrt{r_{var} + \epsilon}} * w^{\mathcal{B}_3^{Res-50}} + b^{\mathcal{B}_3^{Res-50}} \quad (64)$$

Down Sampling Down sampling composes of multiple layers, including: another convolution layer, batch normalization layer, the input is the original input $x^{(iC_1^{Res-50})}$. So we use the convolution layer to define it.

Let filter A^{C^d} as the weight matrix, which is a tensor in $\mathbb{R}^{C^d \times \frac{C^{cC_1^{Res-50}}}{groups} \times kH \times kW}$, And the input is in $\mathbb{R}^{1 \times C^{cC_1^{Res-50}} \times H^{cC_1^{Res-50}} \times W^{cC_1^{Res-50}}}$,

The convolution \mathcal{Cd} is mapping from input $x^{(iC_1^{Res-50})} \in \mathbb{R}^{1 \times C^{cC_1^{Res-50}} \times H^{cC_1^{Res-50}} \times W^{cC_1^{Res-50}}}$ to $x^d \in \mathbb{R}^{1 \times C^d \times H^d \times W^d}$ defined as:

$$x^d \in \mathbb{R}^{1 \times C^d \times H^d \times W^d} := \mathcal{Cd}(x^{(iC_1^{Res-50})}) \quad (65)$$

here we let $A^{C^d} \in \mathbb{R}^{256 \times 64 \times 1 \times 1}$. Let input height and weight be: $H^{cC_1^{Res-50}} = 56, W^{cC_1^{Res-50}} = 56, C^{cC_1^{Res-50}} = 64$, the kernel height and weight can be $kH = 1, kW = 1, sH = 1, sW = 1, dH = 1, dW = 1, padH = 0, padW = 0$, it means $kernelsize = 1, stride = 1$, and $padding = 1$, and $dilation = 1$, and $groups = 1$. $iC_1^{Res-50} = 64$ and $iW^{cC_1^{Res-50}} = iH^{cC_1^{Res-50}} = 56$. And we have $C^d = 256, H^d = W^d = 56$

Same operation as previous batch normalization layer, here a batch normalization \mathcal{Bd} is defined as

$$x^{bd} \in \mathbb{R}^{1 \times C^d \times H^d \times W^d} := \mathcal{Bd}(x^d) = \frac{x^d - r_{mean}^{bd}}{\sqrt{r_{var}^{bd} + \epsilon}} * w^{(bd)} + b^{(bd)} \quad (66)$$

Addition The layer before the last layer is an addition operation, and its second input can be x^{bd} or $x^{(iC_1^{Res-50})}$ depends on if the down-sampling layers exist.

$$x^{(A^{Res-50})} \in \mathbb{R}^{1 \times C^{(C_3^{Ret-50})} \times H^{(C_3^{Ret-50})} \times W^{(C_3^{Ret-50})}} := x^{(B_3^{Res-50})} + x^{bd} \quad (67)$$

ReLU layer - $\mathcal{R}_3^{(Res-50)}$ The rectified liner unit function is applied afterwards such that

$$x^{(R_3^{Res-50})} \in \mathcal{R}^{1 \times C^{(C_3^{Ret-50})} \times H^{(C_3^{Ret-50})} \times W^{(C_3^{Ret-50})}} := R_3^{(Res-50)}(x^{(A^{Res-50})}) : max(0, x^{(A^{Res-50})}) \quad (68)$$

where $C^{(C_3^{Ret-50})} = 256$, $H^{(C_3^{Ret-50})} = 56$ $W^{(C_3^{Ret-50})} = 56$.

Summary of Sequential Layers Here we give a summary of the sequential layers.

The next bottleneck module's input channel size is the previous module's output channel size.

So the $x^{(MP_0^{Res-50})}$ pass through series modules. We give the details in the following Appendix ??.

\mathcal{S}_1 sequential layer 1:

$$x^{(S_1, L_1^d)} \in \mathbb{R}^{1 \times C^{((S_1, L_1^d))} \times H^{(S_1, L_1^d)} \times W^{(S_1, L_1^d)}} := \mathcal{L}_1^d(x^{(MP_0^{Res-50})} \in \mathbb{R}^{1 \times 64 \times 56 \times 56}) \quad (69)$$

where $C^{((S_1, L_1^d))} = 256$, and $H^{(S_1, L_1^d)} = H^{(S_1, L_1^d)} = 56$

In detail,

$$\begin{aligned} x^{(S_1, L_1^d)} &:= \mathcal{L}_1^d(x^{(MP_0^{Res-50})}) = \\ &\circ \mathcal{C}_1^{(S_1, L_1^d, Res-50)} \circ \mathcal{B}_1^{(S_1, L_1^d, Res-50)} \circ Relu_1^{(S_1, L_1^d, Res-50)} \\ &\circ \mathcal{C}_2^{(S_1, L_1^d, Res-50)} \circ \mathcal{B}_2^{(S_1, L_1^d, Res-50)} \circ Relu_2^{(S_1, L_1^d, Res-50)} \\ &\circ \mathcal{C}_3^{(S_1, L_1^d, Res-50)} \circ \mathcal{B}_3^{(S_1, L_1^d, Res-50)} \circ Downsampling^{(S_1, L_1^d, Res-50)} \\ &\circ Addition_1^{(S_1, L_1^d, Res-50)} \circ Relu_3^{(S_1, L_1^d, Res-50)}(x^{(MP_0^{Res-50})}) \end{aligned} \quad (70)$$

And,

$$x^{S_1, L_2^b} \in \mathbb{R}^{1 \times 256 \times 56 \times 56} := \mathcal{L}_2^b(x) \quad (71)$$

In detail,

$$\begin{aligned} x^{(S_1, L_2^b)} &:= \mathcal{L}_2^b(x^{(S_1, L_1^d)}) \in \mathbb{R}^{1 \times 256 \times 56 \times 56} = \\ &\circ \mathcal{C}_1^{(S_1, L_2^b, Res-50)} \circ \mathcal{B}_1^{(S_1, L_2^b, Res-50)} Relu_1^{(S_1, L_2^b, Res-50)} \circ \\ &\circ \mathcal{C}_2^{(S_1, L_2^b, Res-50)} \circ \mathcal{B}_2^{(S_1, L_2^b, Res-50)} \circ Relu_2^{(S_1, L_2^b, Res-50)} \circ \\ &\circ \mathcal{C}_3^{(S_1, L_2^b, Res-50)} \circ \mathcal{B}_3^{(S_1, L_2^b, Res-50)} \circ \\ &\circ Addition_1^{(S_1, L_2^b, Res-50)} \circ Relu_3^{(S_1, L_2^b, Res-50)}(x^{(S_1, L_1^d)}) \end{aligned} \quad (72)$$

$$x^{S_1, L_3^b} \in \mathbb{R}^{1 \times 256 \times 56 \times 56} := \mathcal{L}_3^b(x) \quad (73)$$

In detail,

$$\begin{aligned} x^{(S_1, L_3^b)} &:= \mathcal{L}_3^b(x^{(S_1, L_2^b)}) \in \mathbb{R}^{1 \times 256 \times 56 \times 56} = \\ &\circ \mathcal{C}_1^{(S_1, L_3^b, Res-50)} \circ \mathcal{B}_1^{(S_1, L_3^b, Res-50)} Relu_1^{(S_1, L_3^b, Res-50)} \circ \\ &\circ \mathcal{C}_2^{(S_1, L_3^b, Res-50)} \circ \mathcal{B}_2^{(S_1, L_3^b, Res-50)} \circ Relu_2^{(S_1, L_3^b, Res-50)} \circ \\ &\circ \mathcal{C}_3^{(S_1, L_3^b, Res-50)} \circ \mathcal{B}_3^{(S_1, L_3^b, Res-50)} \circ \\ &\circ Addition_1^{(S_1, L_3^b, Res-50)} \circ Relu_3^{(S_1, L_3^b, Res-50)}(x^{(S_1, L_3^b)}) \end{aligned} \quad (74)$$

\mathcal{S}_2 sequential layer 2:

$$x^{(S_2, L_1^d)} \in \mathbb{R}^{1 \times 256 \times 28 \times 28} := \mathcal{L}_1^d(x_o \in \mathbb{R}^{1 \times 256 \times 56 \times 56}) \quad (75)$$

In detail,

$$\begin{aligned} x^{(S_2, L_1^d)} &:= \mathcal{L}_1^d(x^{(S_1, L_3^b)}) \in \mathbb{R}^{1 \times 256 \times 56 \times 56} = \\ &\circ \mathcal{C}_1^{(S_2, L_1^d, Res-50)} \circ \mathcal{B}_1^{(S_2, L_1^d, Res-50)} \circ Relu_1^{(S_2, L_1^d, Res-50)} \circ \\ &\circ \mathcal{C}_2^{(S_2, L_1^d, Res-50)} \circ \mathcal{B}_2^{(S_2, L_1^d, Res-50)} \circ Relu_2^{(S_2, L_1^d, Res-50)} \circ \\ &\circ \mathcal{C}_3^{(S_2, L_1^d, Res-50)} \circ \mathcal{B}_3^{(S_2, L_1^d, Res-50)} \circ Downsampling^{(S_2, L_1^d, Res-50)} \\ &\circ Addition_1^{(S_2, L_1^d, Res-50)} \circ Relu_3^{(S_2, L_1^d, Res-50)}(x^{(S_1, L_3^b)}) \end{aligned} \quad (76)$$

And,

$$x^{(S_2, L_2^b)} \in \mathbb{R}^{1 \times 512 \times 28 \times 28} := \mathcal{L}_2^b(x) \quad (77)$$

In detail,

$$\begin{aligned} x^{(S_1, L_2^b)} &:= \mathcal{L}_2^b(x^{(S_1, L_1^d)}) \in \mathbb{R}^{1 \times 256 \times 56 \times 56} = \\ &\circ \mathcal{C}_1^{(S_1, L_2^b, Res-50)} \circ \mathcal{B}_1^{(S_1, L_2^b, Res-50)} Relu_1^{(S_1, L_2^b, Res-50)} \circ \\ &\circ \mathcal{C}_2^{(S_1, L_2^b, Res-50)} \circ \mathcal{B}_2^{(S_1, L_2^b, Res-50)} \circ Relu_2^{(S_1, L_2^b, Res-50)} \circ \\ &\circ \mathcal{C}_3^{(S_1, L_2^b, Res-50)} \circ \mathcal{B}_3^{(S_1, L_2^b, Res-50)} \circ \\ &\circ Addition_1^{(S_1, L_2^b, Res-50)} \circ Relu_3^{(S_1, L_2^b, Res-50)}(x^{(S_1, L_1^d)}) \end{aligned} \quad (78)$$

$$x^{(S_2, L_3^b)} \in \mathbb{R}^{1 \times 512 \times 28 \times 28} := \mathcal{L}_3^b(x) = \quad (79)$$

$$\circ \mathcal{C}_1^{(S_2, L_3^b, Res-50)} \circ \mathcal{B}_1^{(S_2, L_3^b, Res-50)} \circ Relu_1^{(S_2, L_3^b, Res-50)} \quad (80)$$

$$\circ \mathcal{C}_2^{(S_2, L_3^b, Res-50)} \circ \mathcal{B}_2^{(S_2, L_3^b, Res-50)} \circ Relu_2^{(S_2, L_3^b, Res-50)}$$

$$\circ \mathcal{C}_3^{(S_2, L_3^b, Res-50)} \circ \mathcal{B}_3^{(S_2, L_3^b, Res-50)} \circ Addition_1^{(S_2, L_3^b, Res-50)} \circ Relu_3^{(S_2, L_3^b, Res-50)}(x^{(S_1, L_2^b)})$$

$$x^{(S_2, L_3^b)} \in \mathbb{R}^{1 \times 512 \times 28 \times 28} := \mathcal{L}_4^b(x) = \quad (81)$$

$$\circ \mathcal{C}_1^{(S_2, L_3^b, Res-50)} \circ \mathcal{B}_1^{(S_2, L_3^b, Res-50)} \circ Relu_1^{(S_2, L_3^b, Res-50)}$$

$$\circ \mathcal{C}_2^{(S_2, L_3^b, Res-50)} \circ \mathcal{B}_2^{(S_2, L_3^b, Res-50)} \circ Relu_2^{(S_2, L_3^b, Res-50)}$$

$$\circ \mathcal{C}_3^{(S_2, L_3^b, Res-50)} \circ \mathcal{B}_3^{(S_2, L_3^b, Res-50)} \circ Addition_1^{(S_2, L_3^b, Res-50)} \circ Relu_3^{(S_2, L_3^b, Res-50)}(x^{(S_2, L_3^b)})$$

\mathcal{S}_3 sequential layer 3:

$$x^{(S_3, L_1^d)} \in \mathbb{R}^{1 \times 1024 \times 14 \times 14} := \mathcal{L}_1^d(x^{(S_2, L_4^b)} \in \mathbb{R}^{1 \times 512 \times 28 \times 28}) \quad (82)$$

In detail,

$$x^{(S_3, L_1^d)} := \mathcal{L}_1^d(x^{(S_2, L_4^b)} \in \mathbb{R}^{1 \times 512 \times 28 \times 28}) = \quad (83)$$

$$\circ \mathcal{C}_1^{(S_3, L_1^d, Res-50)} \circ \mathcal{B}_1^{(S_3, L_1^d, Res-50)} \circ Relu_1^{(S_3, L_1^d, Res-50)}$$

$$\circ \mathcal{C}_2^{(S_3, L_1^d, Res-50)} \circ \mathcal{B}_2^{(S_3, L_1^d, Res-50)} \circ Relu_2^{(S_3, L_1^d, Res-50)}$$

$$\circ \mathcal{C}_3^{(S_3, L_1^d, Res-50)} \circ \mathcal{B}_3^{(S_3, L_1^d, Res-50)} \circ Downsampling^{(S_3, L_1^d, Res-50)}$$

$$\circ Addition_1^{(S_3, L_1^d, Res-50)} \circ Relu_3^{(S_3, L_1^d, Res-50)}(x^{(S_2, L_4^b)})$$

And,

$$x^{(S_3, L_2^b)} \in \mathbb{R}^{1 \times 1024 \times 14 \times 14} := \mathcal{L}_2^b(x^{(S_3, L_1^d)}) \quad (84)$$

In detail,

$$x^{(S_3, L_2^b)} := \mathcal{L}_2^b(x^{(S_3, L_1^d)} \in \mathbb{R}^{1 \times 1024 \times 14 \times 14}) = \quad (85)$$

$$\circ \mathcal{C}_1^{(S_3, L_2^b, Res-50)} \circ \mathcal{B}_1^{(S_3, L_2^b, Res-50)} \circ Relu_1^{(S_3, L_2^b, Res-50)} \circ$$

$$\circ \mathcal{C}_2^{(S_3, L_2^b, Res-50)} \circ \mathcal{B}_2^{(S_3, L_2^b, Res-50)} \circ Relu_2^{(S_3, L_2^b, Res-50)} \circ$$

$$\circ \mathcal{C}_3^{(S_3, L_2^b, Res-50)} \circ \mathcal{B}_3^{(S_3, L_2^b, Res-50)} \circ$$

$$\circ Addition_1^{(S_3, L_2^b, Res-50)} \circ Relu_3^{(S_3, L_2^b, Res-50)}(x^{(S_3, L_1^d)})$$

$$\begin{aligned}
x^{(S_3, L_3^b)} \in \mathbb{R}^{1 \times 1024 \times 14 \times 14} := \mathcal{L}_2^b(x^{(S_3, L_2^b)} \in \mathbb{R}^{1 \times 1024 \times 14 \times 14}) = & \\
C_1^{(S_3, L_3^b, Res-50)} \circ \mathcal{B}_1^{(S_3, L_2^b, Res-50)} Relu_1^{(S_3, L_2^b, Res-50)} \circ & \\
C_2^{(S_3, L_3^b, Res-50)} \circ \mathcal{B}_2^{(S_3, L_2^b, Res-50)} \circ Relu_2^{(S_3, L_2^b, Res-50)} \circ & \\
C_3^{(S_3, L_3^b, Res-50)} \circ \mathcal{B}_3^{(S_3, L_2^b, Res-50)} \circ & \\
Addition_1^{(S_3, L_3^b, Res-50)} \circ Relu_3^{(S_3, L_3^b, Res-50)}(x^{(S_3, L_2^b)}) &
\end{aligned} \tag{86}$$

$$\begin{aligned}
x^{(S_3, L_3^b)} \in \mathbb{R}^{1 \times 1024 \times 14 \times 14} := \mathcal{L}_2^b(x^{(S_3, L_2^b)} \in \mathbb{R}^{1 \times 1024 \times 14 \times 14}) = & \\
C_1^{(S_3, L_3^b, Res-50)} \circ \mathcal{B}_1^{(S_3, L_2^b, Res-50)} Relu_1^{(S_3, L_2^b, Res-50)} \circ & \\
C_2^{(S_3, L_3^b, Res-50)} \circ \mathcal{B}_2^{(S_3, L_2^b, Res-50)} \circ Relu_2^{(S_3, L_2^b, Res-50)} \circ & \\
C_3^{(S_3, L_3^b, Res-50)} \circ \mathcal{B}_3^{(S_3, L_2^b, Res-50)} \circ & \\
Addition_1^{(S_3, L_3^b, Res-50)} \circ Relu_3^{(S_3, L_3^b, Res-50)}(x^{(S_3, L_2^b)}) &
\end{aligned} \tag{87}$$

$$\begin{aligned}
x^{(S_3, L_4^b)} \in \mathbb{R}^{1 \times 1024 \times 14 \times 14} := \mathcal{L}_2^b(x^{(S_3, L_2^b)} \in \mathbb{R}^{1 \times 1024 \times 14 \times 14}) = & \\
C_1^{(S_3, L_3^b, Res-50)} \circ \mathcal{B}_1^{(S_3, L_2^b, Res-50)} Relu_1^{(S_3, L_2^b, Res-50)} \circ & \\
C_2^{(S_3, L_3^b, Res-50)} \circ \mathcal{B}_2^{(S_3, L_2^b, Res-50)} \circ Relu_2^{(S_3, L_2^b, Res-50)} \circ & \\
C_3^{(S_3, L_3^b, Res-50)} \circ \mathcal{B}_3^{(S_3, L_2^b, Res-50)} \circ & \\
Addition_1^{(S_3, L_3^b, Res-50)} \circ Relu_3^{(S_3, L_3^b, Res-50)}(x^{(S_3, L_4^b)}) &
\end{aligned} \tag{88}$$

$$\begin{aligned}
x^{(S_3, L_6^b)} \in \mathbb{R}^{1 \times 1024 \times 14 \times 14} := \mathcal{L}_2^b(x^{(S_3, L_5^b)} \in \mathbb{R}^{1 \times 1024 \times 14 \times 14}) = & \\
C_1^{(S_3, L_3^b, Res-50)} \circ \mathcal{B}_1^{(S_3, L_2^b, Res-50)} Relu_1^{(S_3, L_2^b, Res-50)} \circ & \\
C_2^{(S_3, L_3^b, Res-50)} \circ \mathcal{B}_2^{(S_3, L_2^b, Res-50)} \circ Relu_2^{(S_3, L_2^b, Res-50)} \circ & \\
C_3^{(S_3, L_3^b, Res-50)} \circ \mathcal{B}_3^{(S_3, L_2^b, Res-50)} \circ & \\
Addition_1^{(S_3, L_3^b, Res-50)} \circ Relu_3^{(S_3, L_3^b, Res-50)}(x^{(S_3, L_5^b)}) &
\end{aligned} \tag{89}$$

\mathcal{S}_4 sequential layer 4:

$$x^{(S_4, L_1^d)} \in \mathbb{R}^{1 \times 2048 \times 7 \times 7} := \mathcal{L}_1^d(x^{(S_3, L_6^b)} \in \mathbb{R}^{1 \times 1024 \times 14 \times 14}) \tag{90}$$

In detail,

$$\begin{aligned}
x^{(S_4, L_1^d)} &\in \mathbb{R}^{1 \times 2048 \times 7 \times 7} := \mathcal{L}_1^d(x^{(S_3, L_6^b)}) \in \mathbb{R}^{1 \times 1024 \times 14 \times 14} = \\
&\circ C_1^{(S_3, L_1^d, Res-50)} \circ \mathcal{B}_1^{(S_4, L_1^d, Res-50)} \circ Relu_1^{(S_3, L_1^d, Res-50)} \\
&\circ C_2^{(S_4, L_1^d, Res-50)} \circ \mathcal{B}_2^{(S_4, L_1^d, Res-50)} \circ Relu_2^{(S_3, L_1^d, Res-50)} \\
&\circ C_3^{(S_4, L_1^d, Res-50)} \circ \mathcal{B}_3^{(S_4, L_1^d, Res-50)} \circ \text{Downsampling}^{(S_4, L_1^d, Res-50)} \\
&\circ \text{Addition}_1^{(S_4, L_1^d, Res-50)} \circ Relu_3^{(S_4, L_1^d, Res-50)}(x^{(S_3, L_6^b)})
\end{aligned} \tag{91}$$

And,

$$x^{(S_4, L_2^b)} \in \mathbb{R}^{1 \times 2048 \times 7 \times 7} := \mathcal{L}_2^b(x) \tag{92}$$

In detail,

$$\begin{aligned}
x^{(S_4, L_2^b)} &\in \mathbb{R}^{1 \times 2048 \times 7 \times 7} := \mathcal{L}_2^b(x^{(S_4, L_1^d)}) \in \mathbb{R}^{1 \times 2048 \times 7 \times 7} = \\
&\circ C_1^{(S_3, L_2^b, Res-50)} \circ \mathcal{B}_1^{(S_4, L_2^b, Res-50)} \circ Relu_1^{(S_4, L_2^b, Res-50)} \circ \\
&\circ C_2^{(S_4, L_2^b, Res-50)} \circ \mathcal{B}_2^{(S_4, L_2^b, Res-50)} \circ Relu_2^{(S_4, L_2^b, Res-50)} \circ \\
&\circ C_3^{(S_4, L_2^b, Res-50)} \circ \mathcal{B}_3^{(S_4, L_2^b, Res-50)} \circ \\
&\circ \text{Addition}_1^{(S_4, L_2^b, Res-50)} \circ Relu_3^{(S_4, L_2^b, Res-50)}(x^{(S_4, L_1^d)})
\end{aligned} \tag{93}$$

$$\begin{aligned}
x^{(S_4, L_3^b)} &\in \mathbb{R}^{1 \times 2048 \times 7 \times 7} := \mathcal{L}_3^b(x^{(S_4, L_2^b)}) \in \mathbb{R}^{1 \times 2048 \times 7 \times 7} = \\
&\circ C_1^{(S_3, L_3^b, Res-50)} \circ \mathcal{B}_1^{(S_4, L_3^b, Res-50)} \circ Relu_1^{(S_4, L_3^b, Res-50)} \circ \\
&\circ C_2^{(S_4, L_3^b, Res-50)} \circ \mathcal{B}_2^{(S_4, L_3^b, Res-50)} \circ Relu_2^{(S_4, L_3^b, Res-50)} \circ \\
&\circ C_3^{(S_4, L_3^b, Res-50)} \circ \mathcal{B}_3^{(S_4, L_3^b, Res-50)} \circ \\
&\circ \text{Addition}_1^{(S_4, L_3^b, Res-50)} \circ Relu_3^{(S_4, L_3^b, Res-50)}(x^{(S_4, L_2^b)})
\end{aligned} \tag{94}$$

where $C^{(S_4, L_3^b)} = 2048$, and $H^{(S_4, L_3^b)} = W^{(S_4, L_3^b)} = 7$

$$\begin{aligned}
x^{(S_4, L_4^b)} &\in \mathbb{R}^{1 \times 2048 \times 7 \times 7} := \mathcal{L}_4^b(x^{(S_3, L_3^b)}) \in \mathbb{R}^{1 \times 2048 \times 7 \times 7} = \\
&\circ C_1^{(S_3, L_4^b, Res-50)} \circ \mathcal{B}_1^{(S_4, L_4^b, Res-50)} \circ Relu_1^{(S_4, L_4^b, Res-50)} \circ \\
&\circ C_2^{(S_4, L_4^b, Res-50)} \circ \mathcal{B}_2^{(S_4, L_4^b, Res-50)} \circ Relu_2^{(S_4, L_4^b, Res-50)} \circ \\
&\circ C_3^{(S_4, L_4^b, Res-50)} \circ \mathcal{B}_3^{(S_4, L_4^b, Res-50)} \circ \\
&\circ \text{Addition}_1^{(S_4, L_4^b, Res-50)} \circ Relu_3^{(S_4, L_4^b, Res-50)}(x^{(S_4, L_3^b)})
\end{aligned} \tag{95}$$

where $C^{(S_4, L_4^b)} = 2048$, and $H^{(S_4, L_4^b)} = W^{(S_4, L_4^b)} = 7$

5.4 Adaptive Average Pooling layer

An adaptive average pooling layer will calculate the kernel size based on the input and output size. Here we have input $x^{(S_4, L_2^b)} \in \mathbb{R}^{1 \times C^{((S_4, L_4^b))} \times H^{(S_4, L_4^b)} \times W^{(S_4, L_4^b)}}$, and we have output $x^{(adp)} \in \mathbb{R}^{1 \times C^{(adp)} \times H^{(adp)} \times W^{(adp)}}$. So we have input size as $ips = H^{(S_4, L_4^b)} \times W^{(S_4, L_4^b)}$, and we have output size as $ops = H^{(adp)} \times W^{(adp)}$. We also have the kernel size as $\lfloor \frac{ips+ops-1}{ops} \rfloor$.

An adaptive pooling layer is function mapping from input $x^{(S_4, L_2^b)} \in \mathbb{R}^{1 \times C^{((S_4, L_4^b))} \times H^{(S_4, L_4^b)} \times W^{(S_4, L_4^b)}}$ to $x^{(adp)} \in \mathbb{R}^{1 \times C^{(adp)} \times H^{(adp)} \times W^{(adp)}}$ defined as:

$$x^{(adp)} \in \mathbb{R}^{1 \times C^{(adp)} \times H^{(adp)} \times W^{(adp)}} := AdaptiveAvgPooling(x^{(S_4, L_2^b)}) \quad (96)$$

where

$$H^{(adp)} = \frac{H^{(S_4, L_4^b)} + 2 \times padH - dH \times (kH - 1) - 1}{sH} + 1 \quad (97)$$

$$W^{(adp)} = \frac{W^{(S_4, L_4^b)} + 2 \times padW - dW \times (kW - 1) - 1}{sW} + 1 \quad (98)$$

Here, the $H^{(S_4, L_4^b)} = W^{(S_4, L_4^b)} = 7$, and the $H^{(adp)} = W^{(adp)} = 1$, So the kernel height and weight can be $kH = 49, kW = 49$, it means $kernel size = 49$, we have $C^{(adp)} = 2048, H^{(adp)} = W^{(adp)} = 1$

Output Representation We are interested in this layer, so the output $x^{(rep)}$ will be retrieved from this layer. So the final $x^{(rep)} \in \mathbb{R}^{C^{(adp)}}$

For example, we get some good statistics of $x^{(rep)}$ such that $mean(x^{(rep)}) = 0.42$ and $std(x^{(rep)}) = 0.47$, $max(x^{(rep)}) = 5.04$ and $min(x^{(rep)}) = 0.0$

5.5 A fully Connected layer

We refactor $x^{(adp)} \in \mathbb{R}^{1 \times C^{(adp)} \times 1 \times 1}$ to $x^{(adp)} \in \mathbb{R}^{1 \times C^{(adp)}}$. And a fully connected layer is after this,

A fully connected layer is a linear transformation layer such that

$$x^{out} \in \mathbb{R}^{1 \times |L|} := \mathcal{F}(x^{(adp)}, A, b) = \alpha x^{(adp)} A^T + \beta b \quad (99)$$

where $A \in \mathbb{R}^{|L| \times C^{(adp)}}$ is a weight matrix, and $b \in \mathbb{R}^{1 \times |L|}$ is bias vector, and parameters $\alpha = 1, \beta = 1$

So the final output after resnet is $x^{out} \in \mathbb{R}^{1 \times |L|}$

6 Simulation of GAN and GMM

6.1 Simulation Experiment

Given the model of GAN and CNN representation, we simulate our representation data from GAN, and constructed the corresponding GMM from the following process. We simulate the GAN model based on the bigGAN in [1] and generate three classes of images. We represent our GAN-data using the Resnet-50 [4] for representation. We also

used the VGG-16 [9] and Densennet201 [5] for representation, but due to resource limit, we didn't provide the model details for them, interested readers are welcome to look up their model by reading their original papers. Suppose we have K classes, and for each class, we let x_i is a vector with dimension p , where p depends on the model of the CNN, which represents the one of the layers of CNN. We set the $x_i = x^{(rep)}$ from our model. We have $p = C^{adp} = 2048$. We have total image sample size N for each class. We let m of N images to construct the statistics of GMM, and the rest $n = N - m$ to do representation of GAN and GMM samples, where m is a parameter between $[1, N]$. We let $\bar{x} = \frac{1}{m} \sum_{i=1}^m x_i$, we construct the noise vector $\mathbf{t} = [N(0, 1) \dots N(0, 1)]$.

We let $j = 1 \dots n$, $\{z_j, y_j\}$ represent the GMM sample and GAN sample respectively. Then we use the following calculations to construct our sample.

$$\underbrace{z_j}_{\text{dimension: } 1 \times p} = \bar{x} + \underbrace{\mathbf{t}}_{p \text{ dimension vector}} \left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})^T \right)^{\frac{1}{2}} \quad (100)$$

$$\underbrace{y_j}_{\text{dimension: } 1 \times p} = x_i, i = 1 \dots n \quad (101)$$

We use previous procedure generate the n samples, so the total sample size is Kn . We denote as \mathbf{X} . The Gram matrix $G \in \mathbb{R}^{Kn \times Kn}$ was calculated as below.

$$G = \frac{1}{p} \mathbf{X} \mathbf{X}^T \quad (102)$$

Then eigenvalues of Gram matrix for both the images and GMM were calculated, which is a Kn dimension vector.

6.2 Experiment Settings

In our simulation experiment, we choose $K = 3$. We also choose $N = 500$, based on the CNN models we choose, our p could equal to 2048, 4096, or 1920 based on which CNN model we used. We choose $m = 330$, then $n = 500 - 330 = 170$,

Based on the previous result, \mathbf{X} is a $510 \times p$ dimension matrix., then the eigenvalues will be the 510 dimension vector. We then conducted the following experiments.

6.3 Distribution of Noise Vector z

We change the sampling distribution of noise vector z . We let the z distributed from uniform $U(0, 1)$, $Bernoulli(0, 1)$ with $P = 0.5$, standard normal spherical Gaussian with mean 0 and covariance matrix $\sigma = I$, Gaussian with mean 0 and a covariance matrix with its entry $\sigma_{ij} = 0.5$, and Gaussian with mean 0 and a covariance matrix with its entry $\sigma_{ij} = 1.0$.

Fig 1, Fig 2, Fig 3 show the distribution of eigenvalues of Gram matrix under different distributions of noise vector z , we have seen that the eigenvalue of Gram matrix is invariant to the distribution of input of the Generator.

6.4 Distribution of Noise Vector t

The GMM is constructed in equation 100, where the covariance is constructed to be the same with the Gram matrix of the GAN data. t in equation 100 is noise vector with distribution of $N(0, 1)$. We then investigated the spectral behavior of the data after we change the distribution of t ,

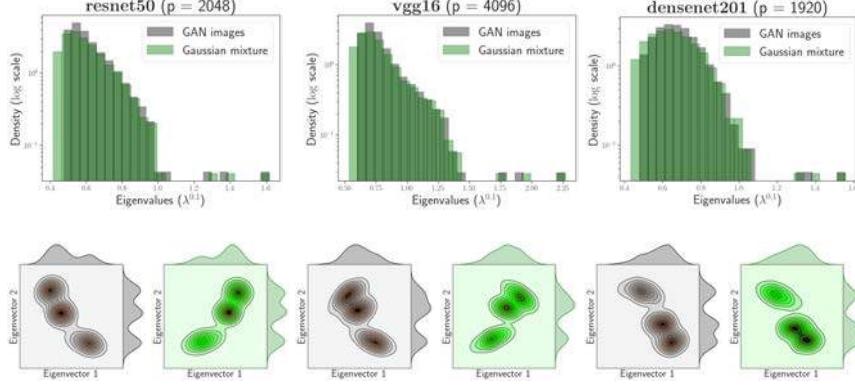


Figure 1: Histogram of The Gram Matrix of GAN Data Given $z \sim U(-1, 1)$ and GMM with Same Mean and Covariance

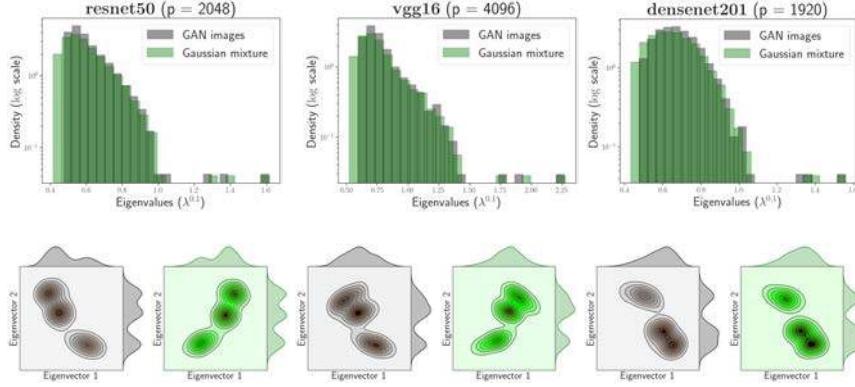


Figure 2: Histogram of The Gram Matrix of GAN Data Given $z \sim Bernoulli(0, 1)$ and GMM with Same Mean and Covariance

We first fix the distribution of $z \sim N(0, 1)$, then change the distribution of t such that

$$t_j = (t_{1j}, \dots, t_{pj}) \in \mathbb{R}^p \sim Bernoulli(-1, 1)$$

. Fig 1 shows the density of eigenvalues of the Gram matrix under such a scenario.

We then change the distribution of $z \sim Bernoulli(0, 1)$, and let the t has distribution

$$t_j = (t_{1j}, \dots, t_{pj}) \in \mathbb{R}^p \sim Bernoulli(-1, 1)$$

. Fig 5 shows the density of eigenvalues of the Gram matrix under this scenario.

Fig 4 and Fig 5 show the distribution of eigenvalues of the Gram matrix still resumes the spectral behavior when the data follows a model rather than standard GMM, where the means remain the same but covariance are different.

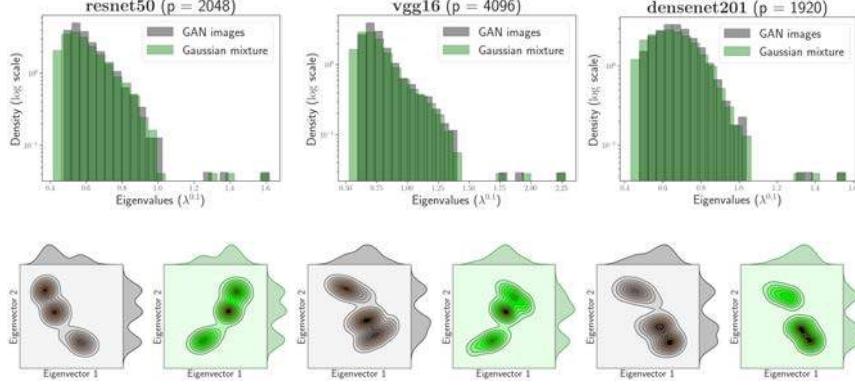


Figure 3: Histogram of The Gram Matrix of GAN Data Given $z \sim N(0, 1)$ and GMM with Same Mean and Covariance

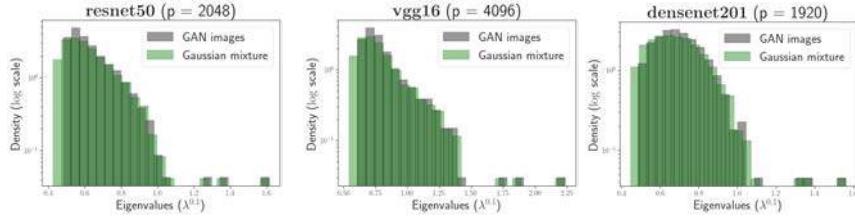


Figure 4: Histogram of The Gram Matrix of GAN Data Given $z \sim N(0, 1)$ and Constructed Model Given $t \sim \text{Bernoulli}(-1, 1)$

7 MNIST Extension

We extend the result into different GAN settings for the MNIST data set.

7.1 Model of MNIST

We constructed a simple MNIST[6] GAN network, where the discriminator has five fully connected hidden layers with leaky ReLU activation, and the generator has to give fully connected layers with leaky ReLU activation followed with batch normalization. The generated images then pass through a pre-trained CNN network for multi-classification. The CNN contains two convolution 2D layers with a max-pooling layer.

7.2 Generation of MNIST Images

We give the generator a noise vector $z \in \mathbb{R}^{100} = N(0, 1)$, and train the neural network with 50 epochs. We sample the generated images during the training, and Fig 6 shows that the generator can generate better images over the training.

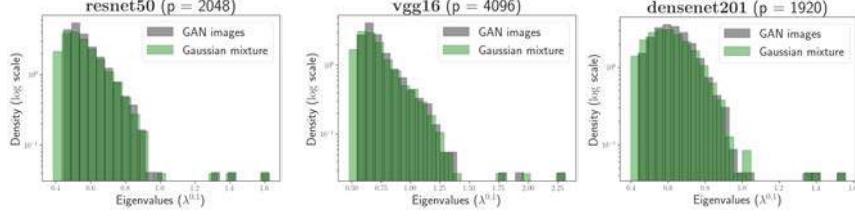


Figure 5: Histogram of The Gram Matrix of GAN Data Given $z \sim \text{Bernoulli}(0, 1)$ and Constructed Model Given $t \sim \text{Bernoulli}(-1, 1)$

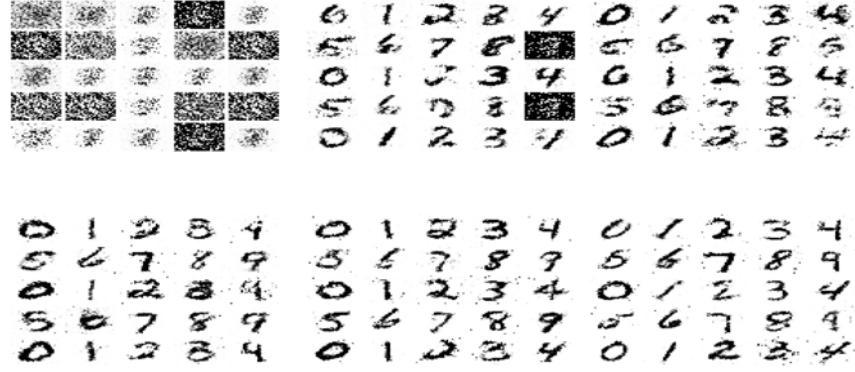


Figure 6: training samples

7.3 Simulation Experiment

In the simulation experiment, we have $k = 10$ classes. We generated $N = 200$ images per class, after passed through CNN, we get the layer output before the second last layer, where the $p = 9216$. We let $n = 50$, so the total matrix $X \in \mathbb{R}^{500 \times 9216}$

Fig 7(a) gives the distribution of the Gram matrix of the above GAN data and corresponding GMM; we see that they have the same distribution.

7.4 Simulation of Multiple Experiments

To further investigate the distribution of the GAN's Gram matrix, we run 500 simulations of the matrix and get the distribution of the second to sixth eigenvalues, and draw the QQ plot and the histogram of them.

Fig 7(b) and Fig 7(c) show the distribution of second to sixth eigenvalues of Gram matrix of the above GAN data over 500 simulations, we can see that they hold a normal distribution.

8 Change The Parameters of Generator

We change the generator's parameters from 10; the original weights are trained based on the Imagenet. However, we give a similar eigenvalue structure of weights by sampling

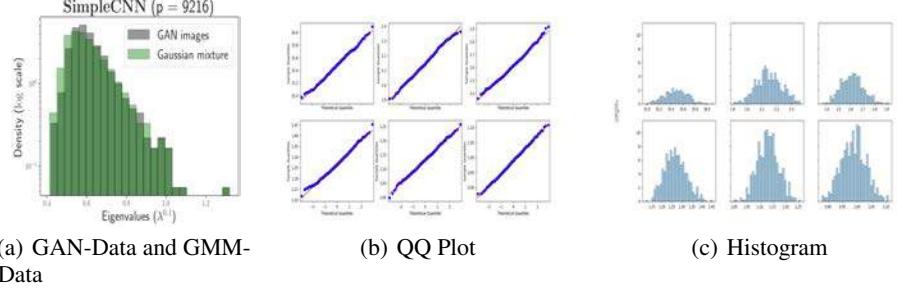


Figure 7: (b) is the QQ plot of 500 simulations of the second to sixth eigenvalues, (c) is the histogram distribution.

the value of the weight from distribution and keeping the relative magnitude of the weights, but not the exact value. We claim that the eigenvalue structure is invariant to the specific weights.

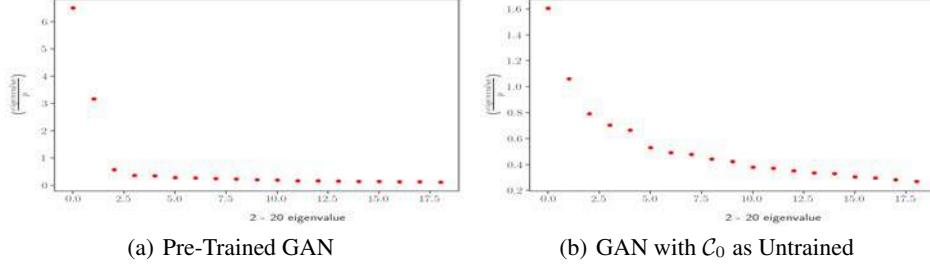


Figure 8: (a) is the second to sixth eigenvalues of Gram matrix of GAN-data with pre-trained GAN, (b) is its counterpart with the convolution layer \mathcal{C}_0 's weights untrained and sampled from a distribution.

Fig 8 indicates that the eigenvalue behavior of the Gram matrix is invariant to the exact GAN weights but is related to the structure of the weights.

References

- [1] Brock, A., Donahue, J., and Simonyan, K. (2019). Large Scale GAN Training for High Fidelity Natural Image Synthesis. *arXiv:1809.11096 [cs, stat]*.
- [2] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee.
- [3] Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative Adversarial Networks. *arXiv:1406.2661 [cs, stat]*.
- [4] He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.
- [5] Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708.
- [6] LeCun, Y., Cortes, C., and Burges, C. (2010). Mnist handwritten digit database.
- [7] Mirza, M. and Osindero, S. (2014). Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*.
- [8] Seddik, M. E. A., Louart, C., Tamaazousti, M., and Couillet, R. (2020). Random Matrix Theory Proves that Deep Learning Representations of GAN-data Behave as Gaussian Mixtures. *arXiv:2001.08370 [cs, stat]*.
- [9] Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

REINFORCEMENT LEARNING UNDER A MULTI-AGENT PREDICTIVE STATE REPRESENTATION MODEL: METHOD AND THEORY

Zhi Zhang, Anonymous authors

ABSTRACT

This paper proposes an optimal policy learning algorithm under a novel multi-agent predictive state representation reinforcement learning model. Compared to the state-of-the-art methods, the most striking feature of our approach is to introduce a dynamic interaction graph to the model, which allows us to represent each agent’s predictive state based on its “neighborhood” agents. Methodologically, we develop an online algorithm that simultaneously learns the predictive state representation and agent policies. Theoretically, we propose an upper bound of the L_2 -norm of the learnt predictive state representation. Empirically, we provide positive numerical results on both a MAMuJoCo robotic learning experiment and a multi-agent particle learning environment to demonstrate the efficacy of the method.

1 INTRODUCTION

Real-world multi-agent systems are considered partially observable since agents do not have a complete perception of the environment states. The predictive state representation (PSR) (Littman et al., 2001) is a representation of the dynamic system state by a vector of predictions of tests conditioned on a history. The tests are sequences of actions and observations, which are true if and only if all the observations occur, given that all the actions taken and the histories are sequences of actions and observations recorded before. It records a sequence of actions and observations and focuses on observable quantities rather than the system’s dynamics. Therefore, compared with learning the hidden-state-based partially observable Markov decision process (POMDP) Kaelbling et al. (1998) models from observation data, learning PSR should be easier and less prone to local minimal problems (Singh et al., 2012). In addition, in the cases where the moment-method is used to learn a model for controlled systems that incorporates actions, PSR is considered as a more general setting to reconstruct the structure of the dynamic system than POMDP (Azizzadenesheli et al., 2016; Hamilton, 2014). Furthermore, (Littman et al., 2001; James & Singh, 2004; Singh et al., 2012) showed that PSR can offer an alternative more compact state representation for POMDP models. Overall, PSR is considered a generalization of POMDP in terms of representation power and the development of learning algorithms.

Due to the above benefits of PSR, it has been used as the basis of some model-based single-agent reinforcement learning (James & Singh, 2004; Boots et al., 2011; Hamilton et al., 2014; Hefny et al., 2018a). Moreover, the predictive states that act as sufficient statistics for the states of multi-agent dynamic systems could potentially help multi-agent decision-making in the complicated, partially observable domains. However, there is no method for developing multi-agent predictive state representation (MAPSR) for multi-agent reinforcement learning (MARL). Due to this reason, we focused on building an algorithm for learning the optimal policies by representing the multi-agent dynamic systems with the MAPSR.

An existing algorithm by (Chen et al., 2020) provides a spectral framework for learning MAPSR using a tensor, with its dimensions representing the agents and joint history. A tensor decomposition is implemented to learn the model parameters. This type of algorithm avoids explicit estimation of the latent state and can get an unbiased estimation if an infinite amount of data is available. However, it is not easy to incorporate prior information such as sparsity or structure because each new source of information leads to new moments and a different and more complex set of equations to solve. (Hefny et al., 2015) represent the distributions in PSR via observable statistics and learn PSR by supervised learning, alleviating the above problem. Later, (Hefny et al., 2018a) extend this method to

develop a new one called predictive state controlled model that can represent the system model where the agents can affect the environment. Due to the trend of deep RL, (Hefny et al., 2018b) developed an end-to-end policy learning method by using the states represented by a predictive state controlled model as the belief state to fit the policy function to train RL. Both the PSR and RL parameters are updated by using a combined loss objective. The PSR part maintains a sufficient representation of distributions under the partially observable environment, the deep RL algorithm brings efficiency for learning.

The environment of MARL can be partially observable and always non-stationary (one agent learning while the other agents’ actions can influence the environment) (Bowling & Veloso, 2002). Predictive state controlled model can model the prediction of systems states affected by actions; if we design a state representation model for MARL environment that can predict the future observations of agents upon exerting future actions of other agents and conditioned on past histories, then the non-stationary environment caused by other agents learning can be modeled. Due to the above reason, we seek to develop a MARL-based PSR model and learning algorithm, which we call MAPSRL. It could potentially ameliorate both the problem of partially observable observations and non-stationarity by other agents’ interaction for multi-agent decision making.

The paper’s first goal is to develop a PSR model for MARL environment. In MARL settings, the distribution of observed observations is dependent on its actions and relevant to other agents and histories. So far, many works (Ryu et al., 2020; Liu et al., 2020b; Niu et al., 2021) that take other agents’ actions into consideration have been shown to achieve competitive performance. In order to consider the interactive relationship under the MAPSR, we introduce a dynamic interaction graph to characterize the “neighborhoods” of agents. Such a graph topology, which is often easy to obtain nowadays, enables us to encode the adjacency relationship of agents into our MAPSR model. We build our model under three common graphs, namely static complete graphs, static non-complete graphs, and dynamic graphs, covering as many real-world scenarios as possible. Under the existence of such a dynamic interaction graph, we build predictions of its observations by considering its neighbor’s actions for every agent. We call this primitive interactive predictions. Then an agent’s PSR will be formulated as a linear combination of those primitive interactive predictions. Due to the decentralized formulation, our model allows implementing the learning of the MAPSR on a single agent level while maintaining other agents’ interactions through the encoding of the interactive graph topology.

As set as our second goal, we analytically upper bound the L_2 -norm difference between the underlying true and learned predictive state representations, providing theoretical guarantees for the performance of our learning algorithm under the MAPSR model in all three graph settings.

Our paper’s third goal is to develop an algorithm that integrates the MAPSR model to the MARL. We develop an online algorithm that simultaneously learns the MAPSR, and the agents’ policies end-to-end with a fully differentiable neural network structure. Our algorithm provides a new model-based MARL, and it has the flexibility to be extended to most of the offline MARL algorithms. We focus on the continuous action space, which is more practical and common to real-world applications such as robotic control.

Finally, we test our algorithm through systematic numerical experiments on MAMujoCo robotic learning experiments (de Witt et al., 2020) and multi-agent particle learning environments (Ryu et al., 2020), and compare our proposed method against two baselines as detailed in Section 6. The results demonstrate the efficiency of our method.

2 RELATED WORKS

Partially Observable Environment. Real-world agents often experience situations that the observed signals are aliased and do not fully determine their state in the system. This is particularly true for multiple agents environments where agents have partial observability due to limited communication (Oliehoek & Amato, 2016). In accommodation to the partially observable environment, POMDP has been adopted by (Kaelbling et al., 1998), and the algorithms (Kaelbling et al., 1998; Cassandra, 1998; Thrun, 1999; Pineau et al., 2003; Poupart & Vlassis, 2008; Platt Jr et al., 2010) for determining an optimal policy have shifted to using the probability distributions (belief state) over the state space instead of exact state space. In general, they have high complexity or suffer from local optima. Moreover, the most common POMDP policy learning assumes the agent has access to a priori knowledge of the system. The access to such prior knowledge has a premise that the agent has

considerable domain knowledge (Kaelbling et al., 1998). However, it is expected that the real-world agents learn the system model and thus a planning policy further without knowledge of the domain.

Overview of PSR. (Littman et al., 2001; James & Singh, 2004) introduced the PSR over an expressive and robust framework for modeling dynamical systems and defined PSR as a representation of state by using a vector of predictions of fully observable quantities (tests) conditioned on past events (histories). A predictive model is constructed directly from execution traces in the PSR framework, utilizing minimal prior information about the domain. The PSR paradigm subsumes POMDP as a special case (Littman et al., 2001). PSR is considered much more compact than POMDP (Aberdeen et al., 2007). The spectral learning method has been proved to show success for learning the PSR (Boots et al., 2011; Jiang et al., 2018). There are other classes of dynamical system learning algorithms that are based on likelihood-based optimization or sampling approaches (Frigola et al., 2013), but they are prone to poor local optima. The spectral learning represents the estimated state by sufficient statistics of future observations and estimates the model parameters by method of moments. However, this line of algorithms is hard to incorporate prior information (Hefny et al., 2015). Thus, (Hefny et al., 2015; 2018a) introduce the supervised learning method to learn PSR and proves its convergence. Although many works study PSR in discrete action space (Hsu et al., 2012; Siddiqi et al., 2010; Boots et al., 2011), (Boots et al., 2013) proposes Hilbert space embedding (HSE)-PSR to deal with continuous actions. (Hefny et al., 2018a) uses an approximation of HSE-PSR by Random Fourier transform (RFF) and built a more principled generalization of PSR to deal with high dimensions. However, all of these studies aim for the single agent scenario.

PSR and RL. The predictive states estimated by the PSR are considered as states in a fully observable Markov Decision Process so that the value function is learned on these states. This line of work has been done in the single-agent environment (Boots & Gordon, 2010; Boots et al., 2011; Hamilton et al., 2014; Venkatraman et al., 2017; Hefny et al., 2018b). Especially, (Hefny et al., 2018b) proposes the recurrent predictive state in the RNN network. Moreover, the learning PSR and policy functions are connected with the end-to-end training.

MAPSR and MARL. MAPSR model is formulated by (Chen et al., 2020) and the model parameters are learnt by tensor decomposition. This formulation is aligned with the spectral algorithm, so it also has the same issue as the single-agent case. MARL algorithms have been developed to learn the Markov Game environment (Lyu & Amato, 2020; Son et al., 2019; Zhang et al., 2019; Rashid et al., 2018; Foerster et al., 2018; Lowe et al., 2017). The partially observable states are pervasive in this environment, bringing non-stationarity to the learning algorithm (Bowling & Veloso, 2002) so that the learning needs to consider the non-stationarity. Recent works propose more sophisticated deep MARL algorithms for multi-agent problems under the paradigm of centralized training with decentralized execution (Zhou et al., 2020; Sunehag et al., 2017; Lowe et al., 2017; Foerster et al., 2018; Rashid et al., 2018). Other than their methods, our approach also accounts for the importance of macroscopic measures of underline systems. Also, providing a predictive state as a belief state rather than the partially observable observations can further help the agent learn. Like the single-agent case, we hope our method can further improve these algorithms.

3 MODEL DESCRIPTION OF MAPSR

3.1 REVIEW OF SINGLE AGENT PSR

Predictive State Representation (PSR). A prediction of a state is defined as the conditional probability of seeing a test’s observations in sequence given that actions of the test are taken in sequence from a history (Littman et al., 2001). Given a finite observation space \mathcal{O} ($o \in \mathcal{O}$) and action space \mathcal{A} ($a \in \mathcal{A}$). A **test** of length k at time t , is defined as a sequence of action-observation pairs that starts at time t and ends at time $t + k - 1$, $\{(o_l, a_l)\}_{l=t}^{t+k-1} = \{o_t, a_t, o_{t+1}, a_{t+1}, \dots, o_l, a_l, \dots, o_{t+k-1}, a_{t+k-1}\}$. A **history**, at time t is a sequence of action-observation pairs that start from the beginning of time and ends at time $t - 1$, $\{(o_l, a_l)\}_{l=1}^{t-1} = \{o_1, a_1, \dots, o_l, a_l, \dots, o_{t-1}, a_{t-1}\}$.

Hilbert Space Embedding Predictive State Controlled Model. In this work, we are interested in extending PSR to decision makings of controlled systems with continuous actions. So we use the model introduced by (Hefny et al., 2018a). In this model, predictive state Q_t satisfies $Q_t \psi_t^a = \mathbb{E}[\psi_t^o | \psi_t^a; \psi_t^h]$ and extended predictive state P_t satisfies $P_t \xi_t^a = \mathbb{E}[\xi_t^o | \xi_t^a; \psi_t^h]$ (i.e., Q_t and P_t are conditional linear expectation operators which maps to the conditional expectation of future observations), where $\psi_t^o := \phi_O(\{o_l\}_{l=t}^{t+k-1})$ and $\psi_t^a := \phi_A(\{a_l\}_{l=t}^{t+k-1})$ are feature maps by kernels k_O, k_A over future observation and action features. The extended predictive state compared

to predictive state adds one more pair of $\{(a_{t+k}, o_{t+k})\}$ to the prediction. The ξ_t^a and ξ_t^o are the corresponding extended feature maps, which satisfy $\xi_t^o = \psi_t^o \otimes \phi_t^o$ and $\xi_t^a = \psi_t^a \otimes \phi_t^a$. Here $\phi_t^o := \phi_o(o_t)$ and $\phi_t^a := \phi_a(a_t)$ are the shorthands for one time feature map by k_o and k_a . We use \otimes to denote the transposed Khatri–Rao product for two matrices with the same number of rows, and each row of the resultant matrix is the vectorized outer product of the corresponding row vectors in the two matrices. Also, we use the $\psi_t^h := \psi^h(\{o_l, a_l\}_{l=1}^{t-1})$ to define a set of features extracted from previous observations and actions (typically from a fixed length window ending at $t - 1$).

3.2 MAPSR FORMULATION

For n -agent systems, we denote joint observations as $\{o_{t,i}\}_{i=1}^n$ and joint actions as $\{a_{t,i}\}_{i=1}^n$ at any time t . The joint test is then a collection of tests for all agents $\{(o_{l,1}, a_{l,1}), \dots, (o_{l,n}, a_{l,n})\}_{l=t}^{t+k-1}$ and the joint history is a collection of histories for all agents $\{(o_{l,1}, a_{l,1}), \dots, (o_{l,n}, a_{l,n})\}_{l=1}^{t-1}$. A prediction of multi-agent state is thus the conditional probability of a joint test given a joint history.

Our Formulation of MAPSR. We bring the HSE-PSR model to the multi-agent settings. For any agent i in an n -agent system, given kernels $\{k_{O_i}\}_{i=1}^n, \{k_{A_i}\}_{i=1}^n, \{k_{o_i}\}_{i=1}^n, \{k_{a_i}\}_{i=1}^n$, we use $\psi_{t,i}^o, \psi_{t,i}^a, \xi_{t,i}^o, \xi_{t,i}^a, \phi_{t,i}^o, \phi_{t,i}^a$ and $\psi_{t,i}^h$ to denote its corresponding feature maps, same as the previous definitions for single agent. The superscripts o, a, h are local observations, actions, and histories for agent i . Moreover, we use the $Q_{t,i}$ and $P_{t,i}$ to denote its predictive state and extended predictive state.

3.3 THE ESTIMATION FOR MAPSR MODEL COMPONENTS

As stated by (Littman et al., 2001), a complete PSR model can build a recursive rule to update itself. In other words, if we are given Q_t , we can get Q_{t+1} via the Bayes’ rules. We use a supervised learning method proposed by (Hefny et al., 2018a) with three types of models to update the predictive states: For any agent i , $P_{t,i} = W_i(Q_{t,i})$, $Q_{t+1,i} = F_i(P_{t,i}, o_{t,i}, a_{t,i})$, and $o_{t,i} = Z_i(Q_{t,i}, a_{t,i})$. Typically, W_i, Z_i are learnable linear maps, and F_i is non-linear but known in advance. Since the model components, $Q_{t,i}$ and $P_{t,i}$ are operators, so they have to be learned as well. The W_i, Z_i can be learned from regression if we know the estimation of $Q_{t,i}$ and $P_{t,i}$, see Appendix A.2.2 for details of estimation of W_i, Z_i . We focus on the details of estimation of $Q_{t,i}, P_{t,i}$ to section 4.1.1.

4 DYNAMIC INTERACTION GRAPH FOR MAPSR MODEL

Many works have considered graph representation of the multi-agent network (Liu et al., 2020b; Ryu et al., 2020). In general, the relationship between agents is characterized as an undirected graph. We introduce a dynamic interaction graph to represent the MAPSR by considering the interaction between agents.

Definition 1. Let $G = (V, E)$, including the set V of nodes and set E of the edges. Each node represents the agent entry, and the edge represents the relationship between the two adjacent agents.

Here we suppose the graph structure is given, that means the number of nodes (n), the number of edges (m), the edge weights, and the maximum number of degrees (k) are available to us. We think that this kind of presupposition is very reasonable because in the real world, for example, there are multiple robots; we can quickly get the geographic position of the robots through sensors, then calculate the structure of the graph formed by them.

4.1 STATIC COMPLETE GRAPH

We starts with a static complete graph G_c , where the relationship between nodes are invariant to time change. A complete graph has $m = n(n - 1)/2$ edges, where $m = |E|$. For each agent, we then represent its PSR by considering other agents’ interactions by

$$Q_{t,i} := g(\{Q_{t,i,j}\}_{j=1}^n) = \sum_j Q_{t,i,j}, \quad (1)$$

same for $P_{t,i}$. To consider the interactive behavior between agents, we introduce two additional notations $Q_{t,i,j}$ and $P_{t,i,j}$. Let $\{(i, j)\}_{i,j=1}^n$ represents a pair of agents on a n multi-agent system. The $Q_{t,i,j}$ is a primitive predictive state of i ’s observation ($\psi_{t,i}^o$) by intervening agent j ’s action ($\psi_{t,j}^a$) and observing agent i ’s observation history and agent j ’s action history ($\psi_{t,i,j}^h$). And $P_{t,i,j}$ is the extended counterpart. If $i == j$, then it becomes an exact single agent scenario. Similarly, we use the same approach to represent $P_{t,i}$. Each agent’s PSR $Q_{t,i}$ and extended PSR $P_{t,i}$ are modeled by fully considering all other available agents.

Based on equation (1), in practice, estimating $Q_{t,i}$ and $P_{t,i}$ denoted as $\hat{Q}_{t,i}$ and $\hat{P}_{t,i}$ requires us to get $\hat{Q}_{t,i,j}$ and $\hat{P}_{t,i,j}$ at first.

4.1.1 ESTIMATION OF $Q_{t,i,j}$, $P_{t,i,j}$

To estimate the P_t and Q_t for single agent, (Hefny et al., 2018a) use the supervised learning method. They show that $\hat{Q}_t = M_{\psi_t^o | \psi_t^a; \psi_t^h} = C_{\psi_t^o \psi_t^a | \psi_t^h} (C_{\psi_t^o \psi_t^a | \phi_t^h} + \lambda I)^{-1}$, where $M_{A|B;c}$ is a linear operator that satisfies $\mathbb{E}[A|B=c] = M_{A|B;c}b$, and $C_{XY} := \mathbb{E}[\phi(X) \otimes \phi(Y)]$ is the uncentered covariance operator, and $C_{XY|z}$ is covariance of X and Y given $Z = z$. They estimate $C_{\psi_t^o \psi_t^a | \psi_t^h}$, $C_{\psi_t^a \psi_t^a | \phi_t^h}$ by sampling data, then get the estimation \hat{Q}_t . The same procedure is used to estimate \hat{P}_t by replacing the features ψ_t^o and ψ_t^a with their extend counterparts ξ_t^o and ξ_t^a .

Similar to the single-agent case above, the representation of $Q_{t,i,j}$ can be achieved as

$$\hat{Q}_{t,i,j} = M_{\psi_{t,i}^o | \psi_{t,j}^a; \psi_{t,i,j}^h} = C_{\psi_{t,i}^o \psi_{t,j}^a | \psi_{t,i,j}^h} (C_{\psi_{t,j}^a \psi_{t,i,j}^h | \phi_{t,i,j}^h} + \lambda I)^{-1} \quad (2)$$

To estimate $C_{\psi_{t,i}^o \psi_{t,j}^a | \psi_{t,i,j}^h}$ and $C_{\psi_{t,j}^a \psi_{t,i,j}^h | \psi_{t,i,j}^h}$, we learn two linear maps $T_{i,j}$ and $U_{i,j}$ such that $T_{i,j}(\psi_{t,i,j}^h) \approx C_{\psi_{t,i}^o \psi_{t,j}^a | \psi_{t,i,j}^h}$ and $U_{i,j}(\psi_{t,i,j}^h) \approx C_{\psi_{t,j}^a \psi_{t,i,j}^h | \psi_{t,i,j}^h}$. The training examples for $T_{i,j}$ and $U_{i,j}$ consist of pairs $(\psi_{t,i,j}^h, \psi_{t,i}^o \otimes \psi_{t,j}^a)$ and $(\psi_{t,i,j}^h, \psi_{t,j}^a \otimes \psi_{t,i}^o)$. The learning of $\hat{P}_{t,i,j}$ can be done in a similar way. More details can be found in Appendix A.2.1

After calculation of $\hat{Q}_{t,i,j}$ and $\hat{P}_{t,i,j}$, under the static complete graph setting where the interaction is considered, defined in equation (1), we can get the estimate of $\hat{Q}_{t,i}$ and $\hat{P}_{t,i}$.

4.1.2 THEORETICAL GUARANTEE OF ESTIMATION OF Q_i , P_i UNDER THE STATIC COMPLETE GRAPH

Theoretically, we show that the difference between Q_i and its estimator \hat{Q}_i is bounded with high probability.

Theorem 1. Let π_Θ be a data collection policy and \mathcal{H} is the range of π_Θ on joint histories. If Equation 1 and 2 used, then for all $h \in \mathcal{H}$ and any $\epsilon \in (0, 1)$, such that $N > \frac{t_{A_j}^2 \log(2d_{A_j}/\epsilon)}{v(C_{\psi_j^a})}$ where N is the number of time points we collect sample, then $\|\hat{Q}_i - Q_i\|$ is bounded as below with probability at least $1 - 3\epsilon$,

$$\|\hat{Q}_i - Q_i\| \leq n\Delta \quad (3)$$

where

$$\Delta = \sqrt{\frac{u(C_{\psi_i^o | \psi_{i,j}^h})}{v(C_{\psi_j^a | \psi_{i,j}^h})^3} \cdot \frac{\|\Delta_1\|^2 + 2u(C_{\psi_j^a | \psi_{i,j}^h})\|\Delta_1\| + \lambda}{v(C_{\psi_j^a | \psi_{i,j}^h})^{(1-\gamma)+\lambda}} + \frac{\|C_{\psi_i^o \psi_j^a | \psi_{i,j}^h}\| \|\Delta_1\| + \|\Delta_2\| \|C_{\psi_j^a | \psi_{i,j}^h}\| + \|\Delta_2\| \|\Delta_1\|}{v(C_{\psi_j^a | \psi_{i,j}^h})^2 (1-\gamma)^2 + \lambda}}.$$

Here Δ_1 , Δ_2 are two other relevant bounds, we provide them in Appendix E.1. $u(\cdot)$, $v(\cdot)$ denote the largest, smallest eigenvalue of a matrix. And $\gamma = \frac{t_{A_j}^2 \log(2d_{A_j}/\epsilon)}{v(C_{\psi_j^a})N} < 1$ is a constant that depends on the magnitude of the norm of ψ_j^a (we assume $\|\psi_j^a\| \leq t_{A_j}$), the dimension of ψ_j^a (d_{A_j}), the (uncentered) covariances ($C_{\psi_j^a} := \mathbb{E}[\psi_j^a \psi_j^{aT}]$) and the sample size (N).

Theorem 1 says we need at least N samples for the bound in equation (3) to be valid. We give the proofs in Appendix E. It is not hard to obtain a bound for P_i using the same approach. We omit that.

4.2 EXTENSION TO STATIC NON-COMPLETE GRAPH

In large-scale multi-agent systems, the number of agents is large, and not all agents need to interact with each other. A static non-completed graph can perfectly represent such a situation. For example, in a given static non-complete graph G_s , we know its maximum number of degree k , and we use the binary $n \times n$ matrix with each entry as $I_{i,j}$ to indicate the interaction between two agents. Then the MAPSR for each agent will be

$$Q_{t,i} := g(\{Q_{t,i,j}\}_{j=1}^n) = \sum_j I_{i,j} Q_{t,i,j}. \quad (4)$$

Lemma 1. Under the same environment depicted in Theorem 1 and given a G_s with k maximum number of the degree to represent agents, then the bound in Theorem 1 can be rewritten as $\|\widehat{Q}_i - Q_i\| \leq k\Delta$.

The conclusion of Lemma 1 is evident since we replace the n with the k neighbors, the total error bound is also decreased approximately as $\frac{k}{n}$.

4.3 DYNAMIC GRAPH

Real-world multi-agents can also formulate a time-dependent dynamic graph G_d , rather than a static graph. A dynamic graph has its structure dynamically changing with time. In other words, the edges can be inserted or deleted across time. The dynamic graph brings more challenges to the representation as the interaction relationship among agents changes constantly.

Braha & Bar-Yam (2009), Ma et al. (2017) and Zhao et al. (2010) consider a dynamic graph as a set of ordered static graphs. For each time point, we are given a static graph such that $G_d = \{G_{d1}, G_{d2}, \dots, G_{dt}\}$, and a time-dependent given binary matrix with $I_{t,i,j}$ indicating the interaction between two agents at each time. Then we have the agent-wise PSR as

$$Q_{t,i} := g(\{Q_{t,i,j}\}_{j=1}^n) = \sum_j I_{t,i,j} Q_{t,i,j}. \quad (5)$$

Compared to (4), the coefficient $I_{t,i,j}$ is time-dependent, which brings a challenge to our theoretical bound. We consider a dynamic graph experiences a trajectory path, assuming every node has a chance at least p to interact by connection with another node at any time point. For example, for a node i , if we take the union set of the nodes interacted with i over the path, then the union set could form a static complete graph; in other words, if i connects j , then we can obtain a valid sample to estimate $Q_{i,j}$ as the complete static graph does, if not, then we skip to the next time point. For the complete static graph, we need the trajectory to run at least N time points to collect enough data to estimate our conditional operator accurately $Q_{i,j}$. Furthermore, The total number of time points needed by i until the N^{th} interaction with j follows a negative binomial distribution $\mathcal{NB}(N, p)$. On average, we need $\frac{N}{p}$ time points before we see i, j completely connecting N times. Now we consider node i could interact with every node in a set of nodes ($J : |J| = n - 1$) for N number of time points. We assume the interaction between two nodes ($i, j \in J$) does not affect their interaction with other nodes. Thus we have a set of independent negative binomial random variables $\{J_l\}_{l=1}^{n-1} \sim \mathcal{NB}(N, p)$ to characterize the interaction of i with $j \in J$. So we are interested in the expectation of the maximum of J_1, \dots, J_{n-1} , a statistics that tells us the expected maximum number of time points of collection of measurements needed for the node i to be able to connect with every node $j \in J$ for at least N number of time points. We denote it as $J_{\{1, \dots, n-1\}}$ and we have

$$\mathbb{E}\{J_{\{1, \dots, n-1\}}\} = \mathbb{E}\{\max(J_1, \dots, J_{n-1})\} = \sum_{N \geq 0} \left(q^N + Npq^{N-1} + \dots + \binom{N}{N-1} p^{N-1} q \right)^{n-1}. \quad (6)$$

Lemma 2. Under the same environment depicted in 1 and given a dynamic graph G_d with every node has a chance at least p to interact with another node in a one-time point, let N be the number of time points we collect data of measurements in order to get the bound in Theorem 1, if Equation 5 and 2 used, then we need at least

$$N' = \left(N - \frac{1}{2} \right) + K(q, n, N) - \frac{\gamma}{\log_{1/q}(1/q)} + F[K(q, n, N)] + \mathcal{O}(1), \quad (7)$$

total number of time points, where $q = 1 - p$, $K(q, n, N) := \log_{1/q}(n - 1) + (N - 1)\log_{1/q}[\log_{1/q}(n - 1)] + (N - 1)\log_{1/q}p - \log_{1/q}(N - 1)!$, F is a periodic C^∞ -function of period 1 and mean value 0 whose Fourier-coefficients are given by $\hat{F}(k) = -\frac{1}{\log(\frac{1}{q})}\Gamma(-\frac{2k\pi i}{\log(\frac{1}{q})})$ for $k \in \mathbb{Z} \setminus \{0\}$. Then $\|\widehat{Q}_i - Q_i\|$ achieves the same bound as in Theorem 1 with probability at least $1 - 3\epsilon$. In other words, $\|\widehat{Q}_i - Q_i\| \leq n\Delta$. (7) is an asymptotic expansion of the right-hand of (6). We give the proof in Appendix E.3.

Lemma 2 gives the worst-case bound for our estimation under the dynamic graph. The result tells us that if we need 1 more sample of measurement for our algorithm to converge on the complete static graph, we need roughly $\log_{1/q} \log_{1/q}(n - 1)$ more samples on the dynamic graph. So as long as we allow enough learning time, the algorithm can converge with high probability.

Complexity with Increasing Agents: For the complete static graph, we need to evaluate $\hat{Q}_{i,j}$ and $\hat{P}_{i,j}$ every time point, which requires $\mathcal{O}(n^2)$ operations and space. Overall, with an increased number of agents, our MAPSR has a polynomial $\mathcal{O}(n^2)$ scaled complexity, which is feasible for learning in a large number of agents environment and is more efficient compared to the centralized MAPSR (Chen et al., 2020) theoretically, which is combinatorially sample complex, the analysis is shown in Appendix A.4. For a non-complete static graph with k maximum number of degrees for $k \ll n$, which is more common in the real world, because a very far-away robot will not likely affect the targeted robot, the operation will be significantly decreased to $\mathcal{O}(k^2)$.

After we get the $\hat{Q}_{t,i}$ and $\hat{P}_{t,i}$, we build the update rule by using W_i, F_i, Z_i , as described in section 3.3.

5 DECISION-MAKING FRAMEWORK WITH MAPSR MODEL

Previous work on single-agent proposed an end-to-end training algorithm (Hefny et al., 2018a) for PSR model and policy learning. Here we design an algorithm for multi-agent settings and incorporate our MARPSR model containing the interactive graph component.

We propose an online learning algorithm to learn the MAPSR and agent policies simultaneously. Our algorithm is shown in Appendix B.2. We call the algorithm **MAPSRL**. The algorithm can be divided into two parts. Firstly, we have a state tracking component, which lets the agents explore the environment and learn an initial predictive state and its parameters W_i and Z_i . The agents first enter the initialization phase (Line 1-4), where we initialize MAPSR by executing an exploration policy to collect data. The agents explore the environment and learn the MAPSR parameters under the given interactive graph. Secondly, we also have a policy component, in which, by starting from the initialized MAPSR and a random policy, the agents experience a T length trajectory under the current policy function that maps predictive states $\{Q_{t,i}\}_{i=1}^n$ to actions $\{a_{t,i}\}_{i=1}^n$ (line 5-18). The agents then update to the next predictive state as introduced in subsection 3.3. Agents also save trajectories (actions, observations, predictive states, next predictive states, and rewards) over the path. Finally, a MARL algorithm conducts the policy learning component. We use a diagram in Appendix B.3 to illustrate this process. We update the MARL and MAPSR model parameter $\Theta = \{\Theta_{\text{MAPSR}}, \Theta_{\text{MARL}}\}$ (line 19) by minimizing the following objective function.

$$\begin{aligned} L(\Theta) &= \alpha_1 l_1(\Theta_{\text{MARL}}) + \alpha_2 l_2(\Theta_{\text{MAPSR}, \text{MARL}}) \\ &= \alpha_1 l_1(\Theta_{\text{MARL}}) + \alpha_2 \frac{1}{n} \sum_{t=0}^T \sum_{i=1}^n E_p(\tau_i | \Theta) \left[\|Z_i(F_i(W_i(g(Q_{t-1,i,j}))) \otimes \psi_{t,i}^a) - \psi_{t,i}^o\|_2^2 \right] \end{aligned} \quad (8)$$

Here the $l_1(\Theta_{\text{MARL}})$ is the loss for MARL, and $l_2(\Theta_{\text{MAPSR}, \text{MARL}})$ is the MSE between prediction and actual observation. And $p(\tau_i | \Theta)$ is the distribution over trajectories induced by the policy and MAPSR. $\Theta_{\text{MAPSR}} = \{W_i, Z_i\}_{i=1}^n$ denotes MAPSR's parameters, and Θ_{MARL} denotes the MARL part. Usually, in a continuous environment, each agent's policy will be parameterized by its actor network that outputs the mean and diagonal covariance of a Gaussian distribution over the continuous action space. We could also give a parameter sharing actor network that maps individual PSR to parameters of a Gaussian distribution over the individual action space if agents are homogeneous.

Now we devise our MAPSR based MARL algorithm in detail. Our algorithm follows the actor-critic framework. We first develop an algorithm under partially observable environments, where each agent has its own critic and actor independently, the gradient of the policy is written as $\nabla_{\theta_i} J(\theta_i) = \mathbb{E}[\nabla_{\theta_i} \log \pi_{\theta_i}(a_i | Q_i)(r_i + \gamma V^{\pi_i}(Q_{t+1,i}) - V^{\pi_i}(Q_i))]$, and the independent critic is updated by minimizing the loss $\mathcal{L} = \mathbb{E}[(V(Q_i) - y_i)^2]$. Here the predictive states Q_i estimated by the MAPSR are considered as states to fit the value and policy functions. We call it **MAPSRL-1**. As we know, independent actor-critic (IAC) (Foerster et al., 2018) would not work well since the environment is not stationary under multi-agent setting, the MADDPG by (Lowe et al., 2017) solves a non-stationary environment by considering other agents actions; however, every agent needs a separate critic that has the global information. Instead, we use a centralized critic like as in COMA (Foerster et al., 2018) and LICA (Zhou et al., 2020) to save space. Our centralized critic minimizes the loss $\mathcal{L} = \mathbb{E}[(V^\pi(Q_1, \dots, Q_n) - y)^2]$. Unlike MADDPG using the deterministic policy, the gradient is taken on the policy parameter space, which is reflected in LICA (Zhou et al., 2020). So our policy gradient is written as $\nabla_{\theta_i} J(\theta_i) = \mathbb{E}[\nabla_{\theta_i} \log \pi_{\theta_i}(a_i | Q_i) \nabla_{\theta_i} V^\pi(Q_1^{\pi_{\theta_1}}, \dots, Q_n^{\pi_{\theta_n}})]$. Also, to solve the non-stationary environments when each agent is learning, it uses the joint predictive states as input. The benefits are 1. PSR (Q) is a function of action (a), bringing agents' action information into the policy gradient, 2. Also, our interactive MAPSR brings the effect of other agents' actions

to the policy gradient, so it does not learn a deterministic policy like MADDPG rather a policy that affects the predictions of future observations considering other agents’ actions 3. Without the need to input other agents’ actions, which saves the input space since the predictive state is a conditional operator that considers other agents’ actions. We call it **MAPSRL-2**. We put our integration details in Appendix B.4.

6 EXPERIMENTS

Environments. We evaluate the performance of our MAPSRL on a collection of MARL tasks under some OpenAI Gym MAMujoco environments (de Witt et al., 2020), such as multi-agent swimmer, hopper, and ant. Each robotic agent is represented as a body graph, where vertices (joints) are connected by adjacent edges (body segments) as shown in Appendix Figure 5. Each agent controls its joints based on the local information observed. All tasks are learned under the partially observable environments by manually hiding some observations for each agent. The goals of the multi-agent systems are aligned with their corresponding single-agent ones. However, different from the single-agent system, the agents in the multi-agent system need to collaborate to reach their goals. For the interactive graph, for simplicity, we considered a complete static graph, where we assume every agent is connected with all other agents in the graph. We defer the details of the robotic agents to Appendix C.1 and the details of the experimental setup to Appendix C.2.

Baselines and Evaluation. We run 50 iterations for each experiment and collect $M = 100$ trajectories in every iteration with a maximum of 1000 steps in every trajectory. After each iteration, we compute the average return $R = \frac{1}{M} \sum_{i=1}^n \sum_{b=1}^M \sum_{t=1}^{T_b} r_{i,b}^t$ on a batch of M trajectories, where, T_b is the length of the b^{th} trajectory. We repeat this process using ten different random seeds and report the average and a standard deviation. To verify the effectiveness of interactive graph, we introduce a baseline called Independent PSR learning (IPSR); in this model, we do not consider the graph, so we formulate n independent single PSR without considering their interactions, which means there are no $Q_{i,j}$ any more but only Q_i . The architecture of MAPSRL-1 and MAPSRL-2 remains the same. To verify the advantage of PSR, we introduce another baseline (MARL) where we take out the MAPSR entirely, so it matches with the MARL run on a partially observable environment. For MAPSRL-1, its MARL baseline is IAC (Foerster et al., 2018), and for MAPSRL-2, is MADDPG (Lowe et al., 2017).

Results And Discussion Figure 1 illustrates the empirical average return vs. the number of interactions with the environment measured in time steps. Our MAPSRL methods consistently outperform IPSR and get the highest rewards under partially observable environments, which justifies the representation power of the interactive graph to assist agents in learning in the non-stationary environment with limited observation when the MARL algorithm has a defect (IAC). Moreover, it can also boost the performance of the existing good MARL algorithm (MADDPG). To further verify the effect of learning the PSR part, we also plotted the predicted trajectory to verify the MAPSR’s performance for predicting the observations in Figure 2. We plotted the predicted observations vs. actual observations in iterations 1, and 40, respectively, for MAPSRL-2. We plotted a row \times columns figure, with each row representing the observation feature and each column representing each agent. As we can see from the figures, the first iteration of the learning does not predict the actual observation very well; it has some mismatches. However, as learning progresses, the predictions

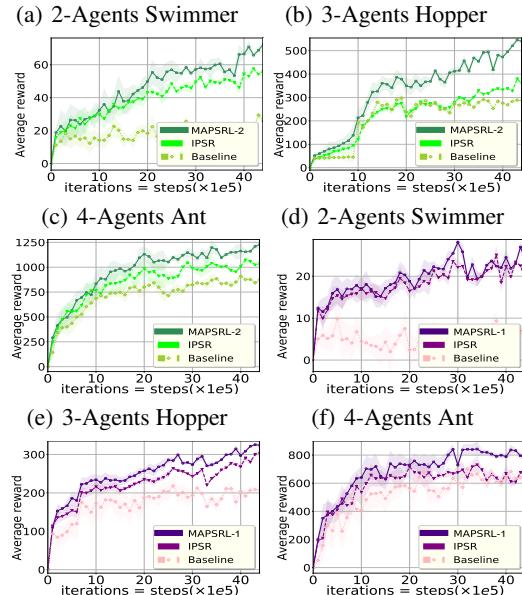


Figure 1: Performance of our method under MA-Mujoco partially observable environments, the ablation study– IPSR, in which we do not consider the agent interaction, and the baseline, which is not using MAPSR. (a)-(c):MAPSRL-2 (It uses centralized critic, and uses gradients of value function with respect to policy parameter, its baseline is MADDPG), (d)-(f):MAPSRL-1 (Its baseline is IAC). We run 10 times and the shaded area is the 95% confidence interval

We repeat this process using ten different random seeds and report the average and a standard deviation. To verify the effectiveness of interactive graph, we introduce a baseline called Independent PSR learning (IPSR); in this model, we do not consider the graph, so we formulate n independent single PSR without considering their interactions, which means there are no $Q_{i,j}$ any more but only Q_i . The architecture of MAPSRL-1 and MAPSRL-2 remains the same. To verify the advantage of PSR, we introduce another baseline (MARL) where we take out the MAPSR entirely, so it matches with the MARL run on a partially observable environment. For MAPSRL-1, its MARL baseline is IAC (Foerster et al., 2018), and for MAPSRL-2, is MADDPG (Lowe et al., 2017).

Results And Discussion Figure 1 illustrates the empirical average return vs. the number of interactions with the environment measured in time steps. Our MAPSRL methods consistently outperform IPSR and get the highest rewards under partially observable environments, which justifies the representation power of the interactive graph to assist agents in learning in the non-stationary environment with limited observation when the MARL algorithm has a defect (IAC). Moreover, it can also boost the performance of the existing good MARL algorithm (MADDPG). To further verify the effect of learning the PSR part, we also plotted the predicted trajectory to verify the MAPSR’s performance for predicting the observations in Figure 2. We plotted the predicted observations vs. actual observations in iterations 1, and 40, respectively, for MAPSRL-2. We plotted a row \times columns figure, with each row representing the observation feature and each column representing each agent. As we can see from the figures, the first iteration of the learning does not predict the actual observation very well; it has some mismatches. However, as learning progresses, the predictions

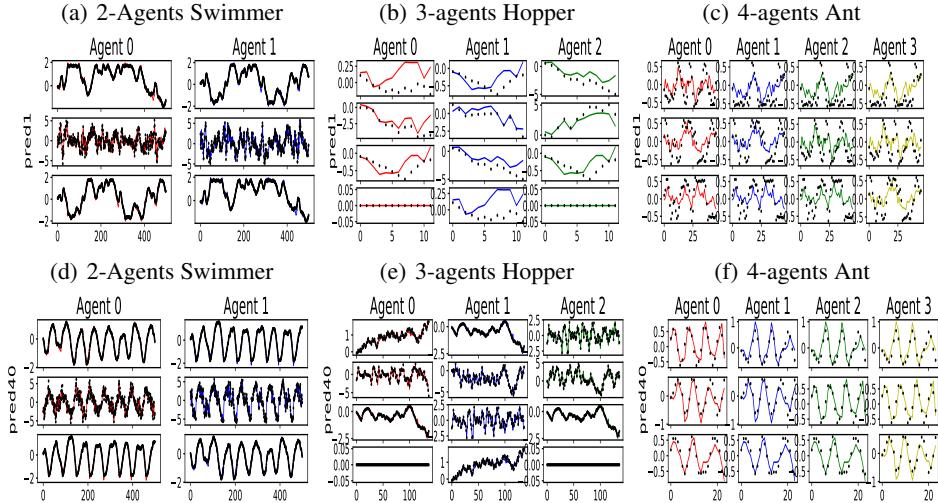


Figure 2: Predicted Trajectories (colored) vs Actual Observations (black). **(a)-(c)** First iteration; **(d)-(f)** Iteration 40. The X axis represents the part of steps encountered for one trajectory under that iteration, and the Y axis represents the numerical value of the observation, i.e. **(a)** has three rows to represent its three coordinates of its observation. We also provide iteration 10, 20 in Appendix D.1

get increasingly more accurate. Note that the actual trajectory is changing according to the current policy, and the current policy is optimized based on the further accurate learning of MAPSR.

To enrich our algorithm environment, we also test our algorithm (MAPSRL-2) in multi-agent particle environments, using the benchmark by (Lowe et al., 2017), please check Appendix D.2 for experiment details. We test our algorithm for large n cases. In this environment, the agents can have cooperative goals such that all agents must maximize a shared return and conflicting competitive goals. We set up the environments where agents can only perform physical actions but not communication; however, to achieve the goals, agents need explicit communication about others' locations to achieve the best reward. These partially observable environments give us the motivation to test our method. We report the rewards in Figure 3. We see that MAPSRL outperforms IPSR and baseline, in terms of the convergence speed and final attained rewards, with a different number of agents. The predictive states convey the information that can help communication between agents in limited communication and observation environments.

7 CONCLUSION

We propose a MAPSR model, extending ideas from single-agent predictive state representations to a multi-agent scenario, during the process, we introduce the dynamic interactive graph to model agents' interactions. Furthermore, we provide the theoretical guarantees of the MAPSR model. Finally, a learning algorithm that supports gradient-based deep MARL methods is developed. Our method provides a model-based MARL framework under a partially observable environment. The experiments proved that our model assumption is valid by observing the highest return while reducing the observations' prediction error over trajectories.

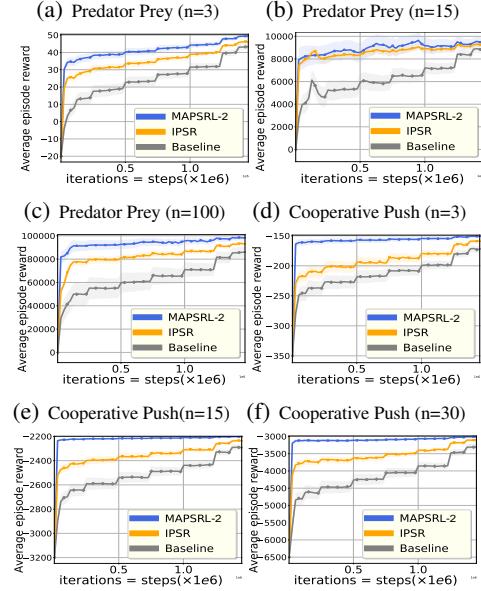


Figure 3: Performance of our method (MAPSRL-2) under multi-agent particle partially observable environments. We use a different number of agents in predator-prey and cooperative push environments. We run ten times, and the shaded area is the 95% confidence interval.

ETHICS STATEMENT

The authors of this paper claimed that we have read the Code of Ethics, adhered to it, and explicitly acknowledged this during the submission process. We claimed that we did not perform any of the following studies that involve human subjects, practices to data set releases, potentially harmful insights, methodologies and applications, potential conflicts of interest and sponsorship, discrimination/bias/fairness concerns, privacy and security issues, legal compliance, and research integrity issues.

REPRODUCIBILITY STATEMENT

We have an appendix section after the main paper to include the model supplements, algorithm details, method diagram, experiment settings, and proofs for our main theorems. We provide a link which directs to our codebase in the supplementary material.

REFERENCES

- Douglas Aberdeen, Olivier Buffet, and Owen Thomas. Policy-gradients for psrs and pomdps. In *Artificial Intelligence and Statistics*, pp. 3–10. PMLR, 2007.
- Kamyar Azizzadenesheli, Alessandro Lazaric, and Animashree Anandkumar. Reinforcement learning of pomdps using spectral methods. In *Conference on Learning Theory*, pp. 193–256. PMLR, 2016.
- Byron Boots and Geoffrey J Gordon. Predictive state temporal difference learning. *arXiv preprint arXiv:1011.0041*, 2010.
- Byron Boots, Sajid M Siddiqi, and Geoffrey J Gordon. Closing the learning-planning loop with predictive state representations. *The International Journal of Robotics Research*, 30(7):954–966, 2011.
- Byron Boots, Geoffrey Gordon, and Arthur Gretton. Hilbert space embeddings of predictive state representations. *arXiv preprint arXiv:1309.6819*, 2013.
- Michael Bowling and Manuela Veloso. Multiagent learning using a variable learning rate. *Artificial Intelligence*, 136(2):215–250, 2002.
- Dan Braha and Yaneer Bar-Yam. Time-dependent complex networks: Dynamic centrality, dynamic motifs, and cycles of social interactions. In *Adaptive Networks*, pp. 39–50. Springer, 2009.
- Anthony Rocco Cassandra. *Exact and approximate algorithms for partially observable Markov decision processes*. Brown University, 1998.
- Bilian Chen, Biyang Ma, Yifeng Zeng, Langcui Cao, and Jing Tang. Tensor Decomposition for Multi-agent Predictive State Representation. *arXiv:2005.13706 [cs]*, May 2020.
- Jakob Foerster, Gregory Farquhar, Triantafyllos Afouras, Nantas Nardelli, and Shimon Whiteson. Counterfactual multi-agent policy gradients. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.
- Roger Frigola, Fredrik Lindsten, Thomas B. Schön, and Carl E. Rasmussen. Bayesian Inference and Learning in Gaussian Process State-Space Models with Particle MCMC. *arXiv:1306.2861 [cs, stat]*, December 2013.
- Peter J Grabner and Helmut Prodinger. Maximum statistics of n random variables distributed by the negative binomial distribution. *Combinatorics, Probability and Computing*, 6(2):179–183, 1997.
- William Hamilton. Compressed predictive state representation: an efficient moment-method for sequence prediction and sequential decision-making. 2014.
- William Hamilton, Mahdi Milani Fard, and Joelle Pineau. Efficient learning and planning with compressed predictive states. *The Journal of Machine Learning Research*, 15(1):3395–3439, 2014.
- Ahmed Hefny, Carlton Downey, and Geoffrey J Gordon. Supervised learning for dynamical system learning. *Advances in neural information processing systems*, 28:1963–1971, 2015.

- Ahmed Hefny, Carlton Downey, and Geoffrey Gordon. An efficient, expressive and local minima-free method for learning controlled dynamical systems. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018a.
- Ahmed Hefny, Zita Marinho, Wen Sun, Siddhartha Srinivasa, and Geoffrey Gordon. Recurrent predictive state policy networks. In *International Conference on Machine Learning*, pp. 1949–1958. PMLR, 2018b.
- Daniel Hsu, Sham M Kakade, and Tong Zhang. A spectral algorithm for learning hidden markov models. *Journal of Computer and System Sciences*, 78(5):1460–1480, 2012.
- Michael R James and Satinder Singh. Learning and discovery of predictive state representations in dynamical systems with reset. In *Proceedings of the twenty-first international conference on Machine learning*, pp. 53, 2004.
- Nan Jiang, Alex Kulesza, and Satinder P Singh. Completing state representations using spectral learning. In *NeurIPS*, pp. 4333–4342, 2018.
- Leslie Pack Kaelbling, Michael L Littman, and Anthony R Cassandra. Planning and acting in partially observable stochastic domains. *Artificial intelligence*, 101(1-2):99–134, 1998.
- Timothy P Lillicrap, Jonathan J Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver, and Daan Wierstra. Continuous control with deep reinforcement learning. *arXiv preprint arXiv:1509.02971*, 2015.
- Michael L Littman, Richard S Sutton, and Satinder P Singh. Predictive representations of state. In *NIPS*, volume 14, pp. 30, 2001.
- Iou-Jen Liu, Raymond A Yeh, and Alexander G Schwing. Pic: permutation invariant critic for multi-agent deep reinforcement learning. In *Conference on Robot Learning*, pp. 590–602. PMLR, 2020a.
- Yong Liu, Weixun Wang, Yujing Hu, Jianye Hao, Xingguo Chen, and Yang Gao. Multi-agent game abstraction via graph attention neural network. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pp. 7211–7218, 2020b.
- Ryan Lowe, Yi Wu, Aviv Tamar, Jean Harb, Pieter Abbeel, and Igor Mordatch. Multi-Agent Actor-Critic for Mixed Cooperative-Competitive Environments. *arXiv:1706.02275 [cs]*, June 2017.
- Xueguang Lyu and Christopher Amato. Likelihood Quantile Networks for Coordinating Multi-Agent Reinforcement Learning. *arXiv:1812.06319 [cs, stat]*, June 2020.
- Shuai Ma, Renjun Hu, Luoshu Wang, Xuelian Lin, and Jinpeng Huai. Fast computation of dense temporal subgraphs. In *2017 IEEE 33rd International Conference on Data Engineering (ICDE)*, pp. 361–372. IEEE, 2017.
- Yaru Niu, Rohan Paleja, and Matthew Gombolay. Multi-agent graph-attention communication and teaming. In *Proceedings of the 20th International Conference on Autonomous Agents and MultiAgent Systems*, pp. 964–973, 2021.
- Frans A Oliehoek and Christopher Amato. *A concise introduction to decentralized POMDPs*. Springer, 2016.
- Judea Pearl et al. Models, reasoning and inference. *Cambridge, UK: Cambridge University Press*, 19, 2000.
- Joelle Pineau, Geoff Gordon, Sebastian Thrun, et al. Point-based value iteration: An anytime algorithm for pomdps. In *IJCAI*, volume 3, pp. 1025–1032. Citeseer, 2003.
- Robert Platt Jr, Russ Tedrake, Leslie Kaelbling, and Tomas Lozano-Perez. Belief space planning assuming maximum likelihood observations. 2010.
- Pascal Poupart and Nikos Vlassis. Model-based bayesian reinforcement learning in partially observable domains. In *Proc Int. Symp. on Artificial Intelligence and Mathematics*, pp. 1–2, 2008.

- Tabish Rashid, Mikayel Samvelyan, Christian Schroeder, Gregory Farquhar, Jakob Foerster, and Shimon Whiteson. Qmix: Monotonic value function factorisation for deep multi-agent reinforcement learning. In *International Conference on Machine Learning*, pp. 4295–4304. PMLR, 2018.
- Heechang Ryu, Hayong Shin, and Jinkyoo Park. Multi-agent actor-critic with hierarchical graph attention network. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pp. 7236–7243, 2020.
- Sajid Siddiqi, Byron Boots, and Geoffrey Gordon. Reduced-Rank Hidden Markov Models. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, pp. 741–748. JMLR Workshop and Conference Proceedings, March 2010.
- Satinder Singh, Michael James, and Matthew Rudary. Predictive state representations: A new theory for modeling dynamical systems. *arXiv preprint arXiv:1207.4167*, 2012.
- Kyunghwan Son, Daewoo Kim, Wan Ju Kang, David Earl Hostallero, and Yung Yi. Qtran: Learning to factorize with transformation for cooperative multi-agent reinforcement learning. In *International Conference on Machine Learning*, pp. 5887–5896. PMLR, 2019.
- James H Stock, Mark W Watson, et al. *Introduction to econometrics*, volume 3. Pearson New York, 2012.
- Peter Sunehag, Guy Lever, Audrunas Gruslys, Wojciech Marian Czarnecki, Vinicius Zambaldi, Max Jaderberg, Marc Lanctot, Nicolas Sonnerat, Joel Z. Leibo, Karl Tuyls, and Thore Graepel. Value-Decomposition Networks For Cooperative Multi-Agent Learning. *arXiv:1706.05296 [cs]*, June 2017.
- Ming Tan. Multi-agent reinforcement learning: Independent vs. cooperative agents. In *Proceedings of the tenth international conference on machine learning*, pp. 330–337, 1993.
- Sebastian Thrun. Monte carlo pomdps. In *NIPS*, volume 12, pp. 1064–1070, 1999.
- Joel A. Tropp. An Introduction to Matrix Concentration Inequalities. *arXiv:1501.01571 [cs, math, stat]*, January 2015.
- Arun Venkatraman, Nicholas Rhinehart, Wen Sun, Lerrel Pinto, Martial Hebert, Byron Boots, Kris M Kitani, and J Andrew Bagnell. Predictive-state decoders: Encoding the future into recurrent networks. *arXiv preprint arXiv:1709.08520*, 2017.
- Christian Schroeder de Witt, Bei Peng, Pierre-Alexandre Kamienny, Philip Torr, Wendelin Böhmer, and Shimon Whiteson. Deep Multi-Agent Reinforcement Learning for Decentralized Continuous Cooperative Control. *arXiv:2003.06709 [cs, stat]*, December 2020.
- Zhi Zhang, Jiachen Yang, and Hongyuan Zha. Integrating independent and centralized multi-agent reinforcement learning for traffic signal network optimization. *arXiv:1909.10651 [cs, stat]*, September 2019.
- Qiankun Zhao, Yuan Tian, Qi He, Nuria Oliver, Ruoming Jin, and Wang-Chien Lee. Communication motifs: a tool to characterize social communications. In *Proceedings of the 19th ACM international conference on Information and knowledge management*, pp. 1645–1648, 2010.
- Meng Zhou, Ziyu Liu, Pengwei Sui, Yixuan Li, and Yuk Ying Chung. Learning Implicit Credit Assignment for Cooperative Multi-Agent Reinforcement Learning. *arXiv:2007.02529 [cs, stat]*, October 2020.

APPENDICES TO REINFORCEMENT LEARNING UNDER A MULTI-AGENT PREDICTIVE STATE REPRESENTATION MODEL: METHOD AND THEORY

A METHOD DETAILS

A.1 MATHEMATICAL MODEL OF PSR

Here, we give the mathematical model for PSR. As in section 3.1, we let $\tau_t := \{(o_l, a_l)\}_{l=t}^{t+k-1}$ as a test (a sequence of action-observation pairs) at time t , then we have a subset $\tau_t^o := \{o_l\}_{l=t}^{t+k-1}$ for observations, and $\tau_t^a := \{a_l\}_{l=t}^{t+k-1}$ for actions. And we let $h_t := \{(o_l, a_l)\}_{l=1}^{t-1}$ as the history at time t . A test is executed at time t if we intervene to select the sequence of actions specified by the test. It is said to succeed at time t if it is executed and the sequence of observations in the test matches the observations by the system (Boots et al., 2013).

The prediction of a length- k test τ_t is defined as the probability distribution that the test succeed at time t , given history h_t : $\Pr(\tau_t^o | \tau_t^a, h_t) = \frac{\Pr(\tau_t^o, \tau_t^a | h_t)}{\Pr(\tau_t^a | h_t)}$

Let $T = \{\tau_z\}, z \in N$ is a finite set of core tests. And we use $T^o = \{\tau_1^o, \dots, \tau_{|T|}^o\}$ and $T^a = \{\tau_1^a, \dots, \tau_{|T|}^a\}$ to indicate its observation and action parts.

A.1.1 LINEAR PSR

A linear PSR is the prediction vector

$$Q_t := \left[\Pr(\tau_1^o | \tau_1^a, h_t), \dots, \Pr(\tau_{|T|}^o | \tau_{|T|}^a, h_t) \right]^\top \quad (9)$$

that contains the probabilities of success of the tests in T , if and only if for any test τ

$$\Pr(\tau^o | \tau^a, h_t) = f_\tau(Q_t), \quad (10)$$

Where $f_\tau : [0, 1]^{|T|} \mapsto [0, 1]$ is a linear projection function. For simplicity of notation, we use the same notations for $Q_t P_t, W, F$, which only follows the meaning defined in this section.

It means that knowing the probabilities for the tests in T is sufficient for computing the probabilities for all other tests in the system. The prediction vector is a sufficient statistic for the system at time t , so we call it as **state** for the **PSR** at time t . Therefore, PSR can represent state by using a vector of predictions of fully observable quantities (tests) conditioned on past events (histories). The prediction function f_τ is linear and has one-to-one relationship to a test τ such that $f_\tau(Q_t) = f_\tau^\top Q_t \quad \forall t$. The linear PSR can still represent systems with nonlinear dynamics.

To maintain predictions in T , we need to update the state Q_t . To do that, we predict the success of any core test τ_z prepended by a new action a and observation o at time $t + k - 1$, which we call $oa\tau_z$.

Based on Bayes rule, we have

$$\Pr(\tau_{z,t+1}^o | \tau_{z,t+1}^a, h_{t+1} = (h_t, o, a)) = \frac{\Pr(o\tau_{z,t+1}^o | a\tau_{z,t+1}^a, h_t)}{\Pr(o | a, h_t)} = \frac{f_{ao\tau_z}^\top Q_t}{f_{ao}^\top Q_t} \quad (11)$$

$f_{ao\tau_z}, f_{ao} \in \mathbb{R}^{|T|}$ are linear operators such that ($\forall \tau_z \in T, \forall a \in \mathcal{A}, \forall o \in \mathcal{O}$). Then let $F_{ao\tau}$ be the matrix with its columns as $f_{ao\tau_z}$ for all $\tau_z \in T$. Then the updated state in PSR is obtained by

$$Q_{t+1} = \frac{F_{ao\tau_z}^\top Q_t}{f_{ao}^\top Q_t} \quad (12)$$

Given the initial prediction vector Q_1 , the PSR can update with equation 12. This recursive application of Bayes rule to a belief state is called the Bayes filter. Now we have seen the extended prediction vector, so we define **extended predictive state** as:

$$P_t := \left[\Pr(a\tau_1^o | a\tau_1^a, h_t), \dots, \Pr(a\tau_{|T|}^o | a\tau_{|T|}^a, h_t) \right]^\top \quad (13)$$

Clearly, we can see there exists a linear extension operator W such that $P_t = W(Q_t)$. And given the new observation, we can have a Bayes filter F such that $Q_{t+1} = F(P_t, o, a)$. We also introduced these operators in section 3.3.

A.1.2 REPRESENTATION OF STATE AS CONDITIONAL EXPECTATION OF SUFFICIENT STATISTICS

Instead of learning the distribution, here we recover the idea of representation of states as a conditional expectation of sufficient statistics and using the supervised learning method to learn (Hefny et al., 2015). Let history h_t^o is a sequence of observation that starts from the beginning of time and ends at time $t - 1$.

We define the belief state $b_t = \Pr(s_t|h_t^o)$, where s_t is the current state of the world. $b_{t+1} = \Pr(s_{t+1}|h_{t+1}^o)$ is the next time belief state. We call b_t as "belief state", which represents the probability distribution over state space, also represents the knowledge and uncertainty about the true state of the system. In dynamic system, the task of getting the updated b_{t+1} with given b_t and new observation o_t is called filtering. The task of estimating the $\Pr(s_{t+1}|h_t^o)$ with given current b_t without incorporating any new observation is called one-step prediction.

Instead maintaining a belief b_t over states, spectral algorithms try to recover observable operators that can be used to perform filtering and prediction directly, by maintaining the expected value of a sufficient statistic of future observations.

Let a test τ_t^o is a length k sequence of observation defined before. As the recursive Bayes rule holds, we can get the next time prediction vector $\Pr(\tau_{t+1}^o|h_{t+1}^o)$ using the new observation o_{t+k} and the current prediction $\Pr(\tau_t^o|h_t^o)$, we also define an extended prediction vector $\Pr(\tau_{t+1}^o|h_t^o)$. These prediction vectors characterize the state of the system and they can be estimated by observable quantities. Given $\Pr(\tau_{t+1}^o|h_t^o)$, filtering becomes the task of getting the updated prediction vector $\Pr(\tau_{t+1}^o|h_{t+1}^o)$, "conditioning" on o_t . One-step prediction becomes getting the $\Pr(\tau_{t+1}^o|h_t^o)$, "marginalizing" over o_t .

Therefore, the spectral algorithms avoid explicitly estimating the latent state or the initial, transition, or observation distributions. We let $Q_t = \mathbb{E}[\nu_t|h_t^o]$, where ν_t a vector of features that determines the distribution of future observations $\Pr(\tau_t^o|h_t^o)$. For simplicity of notation, we use the same notations for Q_t, P_t, W, F . Let $P_t = \mathbb{E}[\varphi_t|h_t^o]$, where φ_t is a vector of features that determines the distribution of observations $\Pr(\tau_t^o|h_t^o)$. We call Q_t is the transformed predictive state. So $Q_{t+1} = \mathbb{E}[\nu_{t+1}|h_{t+1}^o]$, is the updated predictive state, and P_t is the extended predictive state. Let h_t^o be the feature vector of history h_t^o .

In Hidden Markov Models and Kalman filters, the extended state P_t is linearly related to the predictive state Q_t . Which is $P_t = WQ_t$. Estimation of the W can be done using linear regression with samples ν_t and φ_t , however, due to the overlap between observation windows, the noise terms on ν_t and φ_t are correlated, which will cause biased estimate. The instrumental regression (Pearl et al., 2000; Stock et al., 2012) is employed. h_t^o is a instrumental variable that do not overlap with sequence $\{l\}_{l=t}^{l=t+k-1}$ and $\{l\}_{l=t}^{l=t+k}$. The correlation $\text{corr}(h_t^o, e(\varphi_t)) = 0$ and $\text{corr}(h_t^o, e(\nu_t)) = 0$, where e is a measure of error. By taking the conditional expectation of $P_t = WQ_t$ given h_t^o , we have

$$\begin{aligned}\mathbb{E}[P_t|h_t^o] &= \mathbb{E}[WQ_t|h_t^o] \\ \mathbb{E}[\mathbb{E}[\varphi_t|h_t^o]|h_t^o] &= W\mathbb{E}[\mathbb{E}[\nu_t|h_t^o]|h_t^o] \\ \mathbb{E}[\varphi_t|h_t^o] &= W\mathbb{E}[Q_t|h_t^o]\end{aligned}\tag{14}$$

Based on the above relationship, we first estimate the $\mathbb{E}[\varphi_t|h_t^o]$ and $\mathbb{E}[Q_t|h_t^o]$ by sample h_t^o, ν_t , and φ_t , we then use the estimates to compute W . So if we start with Q_1 , we can compute $P_1 = WQ_1$, and get the $Q_{t+1} = F(P_1, o_1)$, where F is the Bayes filter to update the state.

A.2 LEARNING FOR MAPSR IN DETAILS

Here we introduce the details about learning MAPSR, which supplements the section 3.3 and 4.1.1.

A.2.1 LEARNING FOR $\hat{Q}_{t,i,j}$ AND $\hat{P}_{t,i,j}$

To calculate $\hat{Q}_{t,i,j}$, we introduce two sets of linear operators $T_{i,j}$ and $U_{i,j}$, such that $T_{i,j}(\psi_{t,i,j}^h) \approx C_{\psi_{t,i}^o \psi_{t,j}^a | \psi_{t,i,j}^h}$ and $U_{i,j}(\psi_{t,i,j}^h) \approx C_{\psi_{t,j}^a \psi_{t,i}^a | \psi_{t,i,j}^h}$. We estimate them by using two ridge regressions:

$$\arg \min_{T_{i,j}} \sum_{t=1}^T \mathcal{L}(T_{i,j}(\psi_{t,i,j}^h), \psi_{t,i}^o \otimes \psi_{t,j}^a) + R(T_{i,j})\tag{15}$$

$$\arg \min_{U_{i,j}} \sum_{t=1}^T \mathcal{L}(U_{i,j}(\psi_{t,i,j}^h), \psi_{t,j}^a \otimes \psi_{t,j}^a) + R(U_{i,j}) \quad (16)$$

\mathcal{L} represents the ridge regression loss, and R represents the regularizer. After learning $T_{i,j}$ and $U_{i,j}$, we can get the estimate of $C_{\psi_{t,i}^o \psi_{t,j}^a | \psi_{t,i,j}^h}$ and $C_{\psi_{t,j}^a \psi_{t,j}^a | \psi_{t,i,j}^h}$. Then we get the $\hat{Q}_{t,i,j}$ by equation 2. Similarly, we use extended features to obtain $\hat{P}_{t,i,j}$.

A.2.2 EXTENSION, FILTERING AND PREDICTION FUNCTIONS

Filtering. To obtain $\hat{Q}_{t+1,i}$ from $\hat{Q}_{t,i}$, $\hat{P}_{t,i}$, we use filtering. We denote F_i as the filtering function. We describe the filtering process as below. From $\{o_i, a_i\}_{i=1}^n$, we obtain the embedding $\{\phi_{t,i}^o, \phi_{t,i}^a\}_{i=1}^n$. We then compute the observation covariance

$$C_{o_{t,i}, o_{t,i} | h_{t,i}, a_{t,i}} = M_{\phi_{t,i}^o \otimes \phi_{t,i}^o | \phi_{t,i}^a; \psi_{t,i}^h} \phi_{t,i}^a. \quad (17)$$

We then multiply the extended state by inverse observation covariance to change predicting $\phi_{t,i}^o$ into conditioning on $\phi_{t,i}^o$.

$$M_{\psi_{t+1,i}^o | \psi_{t+1,i}^a, \phi_{t,i}^o, \phi_{t,i}^a; \psi_{t,i}^h} = M_{\psi_{t+1,i}^o \otimes \phi_{t,i}^o | \psi_{t+1,i}^a, \phi_{t,i}^a; \psi_{t,i}^h} \times_{\phi_{t,i}^o} (C_{o_{t,i}, o_{t,i} | h_{t,i}, a_{t,i}} + \lambda I)^{-1}. \quad (18)$$

\times here is to denote $n - mode$ (matrix or vector) product, $\times_{\phi_{t,i}^o}$ means multiplying the tensor by a matrix (or vector) in mode $\phi_{t,i}^o$.

We condition on $\phi_{t,i}^o$ and $\phi_{t,i}^a$ to obtain shifted state.

$$Q_{t+1,i} := M_{\psi_{t+1,i}^o | \psi_{t+1,i}^a, \phi_{t,i}^o, \phi_{t,i}^a; \psi_{t,i}^h} = M_{\psi_{t+1,i}^o | \psi_{t+1,i}^a, \phi_{t,i}^o, \phi_{t,i}^a; \psi_{t,i}^h} \times_{\phi_{t,i}^o} \phi_{t,i}^o \times_{\phi_{t,i}^a} \phi_{t,i}^a. \quad (19)$$

Based on the updating rule, $Q_{t+1,i} = F_i(P_{t,i}, o_{t,i}, a_{t,i})$, and $P_{t,i} = W_i(Q_{t,i})$, we write the filtering equation.

$$Q_{t+1,i} = F_i(P_{t,i}, o_{t,i}, a_{t,i}) \\ := M_{\psi_{t+1,i}^o \otimes \phi_{t,i}^o | \psi_{t+1,i}^a, \phi_{t,i}^a; \psi_{t,i}^h} \times_{\phi_{t,i}^o} (M_{\phi_{t,i}^o \otimes \phi_{t,i}^o | \phi_{t,i}^a; \psi_{t,i}^h} \phi_{t,i}^a + \lambda I)^{-1} \times_{\phi_{t,i}^o} \phi_{t,i}^o \times_{\phi_{t,i}^a} \phi_{t,i}^a,$$

where $P_{t,i} := M_{\psi_{t+1,i}^o \otimes \phi_{t,i}^o | \psi_{t+1,i}^a, \phi_{t,i}^a; \psi_{t,i}^h}$. F_i usually is known because it is obtained through the above calculation using known quantities, however, W_i , Z_i must have to be learned by using regressions.

Extension. The W_i can be learned by kernel regression if we know the $\hat{Q}_{t,i}$ and $\hat{P}_{t,i}$. Previous work (Hefny et al., 2018a) demonstrated the kernel regression model for learning single-agent PSR, here we extend to the MAPSR. We set the model parameter W_i .

We optimized a ridge regression problem for W_i .

$$\arg \min_{W_i} \sum_{t=1}^T \mathcal{L}(W_i(\hat{Q}_{t,i}), \hat{P}_{t,i}) + R(W_i)$$

Prediction. We can also get the prediction about the next one time observation o_t by the regression function such as:

$$\hat{o}_{t,i} := \mathbb{E}(o_{t,i} | Q_{t,i}, a_{t,i}) = Z_i(Q_{t,i} \otimes \psi_{t,i}^a)$$

We solve the prediction regression function Z_i by another ridge regression:

$$\arg \min_{Z_i} \sum_{t=1}^T \mathcal{L}(Z_i(\hat{Q}_{t,i} \otimes \psi_{t,i}^a), \psi_{t,i}^o) + R(Z_i)$$

A.3 TENSOR DECOMPOSITION OF MAPSR

Existed Formulation. (Chen et al., 2020) use a $n + 1$ multi-dimensional tensor called system dynamics tensor $\mathcal{D} \in \mathbb{R}^{|\mathcal{T}_1| \times \dots \times |\mathcal{T}_n| \times |\mathcal{H}|}$ to represent the system dynamics of the MAPSR, with $|\mathcal{T}_1|$ representing the cardinality of the tests for agent 1, n denoting the number of agents, and the $|\mathcal{H}|$ being the cardinality of the joint history. Each element of the tensor is a probability of a joint test given joint histories. Given the system dynamics tensor, finding the latent predictive state can be

transferred into finding a minimal linearly independent set from the system dynamics tensor, and it can be solved by spectral method such as tensor decomposition.

As mentioned earlier, this formulation can not satisfy our needs. In **Proposition 1** of Appendix A.4, we also give the sample complexity analysis such that the sample size needed to formulate \mathcal{D} scales exponentially with the length of tests and number of agents.

Here we give a summary of tensor decomposition of their method. Given a system dynamic tensor \mathcal{D} such that:

$$\mathcal{D} \approx [\lambda; D^1, \dots, D^n, F] = \sum_{r=1}^R \lambda_r D_r^1 \circ \dots \circ D_r^n F_r \quad (20)$$

Here \circ is the outer product. D^1, \dots, D^n, F are matrices. The factor matrices D^1, D^n, F consist of the vectors, i.e., $D^1 = [D_{:,1}^1 D_{:,2}^1 \dots D_{:,R}^1] \in \mathbb{R}^{|\mathcal{T}_1| \times R}$. A colon is used to indicate all elements of a mode, thus, the R^{th} column of D^1 is denoted by $D_{:,R}^1$. For any $i_1 \in \{1, \dots, |\mathcal{T}_1|\}, i_n \in \{1, \dots, |\mathcal{T}_n|\}$, and $k \in \{1, \dots, |\mathcal{H}|\}$, after the decomposition, we get

$$\mathcal{D}_{i_1, \dots, i_n, k} = \sum_{r=1}^R \lambda_r D_{i_1 r}^1 \cdots D_{i_n r}^n F_{k r}, \quad (21)$$

where $\{i_1, \dots, i_n, k\}$ are index corresponding to the specific dimension of the \mathcal{D} . The last dimension of the tensor \mathcal{D} is compressed in a matrix F , and its row vector $x_k = [x_k(1) \dots x_k(R)] \in \mathbb{R}^{1 \times R}, k \in \{1, \dots, |\mathcal{H}|\}$ is a summary of joint history and can be considered as a compressed version of the system predictive state vector $p(\mathbf{Q}|\mathbf{h}_k)$, the joint history $\mathbf{h}_k \in \mathcal{H}(k \in [1, |\mathcal{H}|])$ at time step $s = |\mathbf{h}_k|$, where \mathbf{Q} is the core joint test set. The whole fibers listed in the set \mathbf{Q} form a basis of the space spanned by the mode-(n+1) fibers of tensor \mathcal{D} . Thus, by constructing the vector $m = (\lambda * D_{i_1:}^1 * \dots * D_{i_n:}^n)^T$, where $*$ is Hadamard product. $D_{i_1:}^1$ denotes the i_1 -th row vector of D^1 . Then we could rewrite the previous equation as

$$\mathcal{D}_{i_1, \dots, i_n, k} = \sum_{r=1}^R m(r) x_k(r) = x_k (\lambda * D_{i_1:}^1 * \dots * D_{i_n:}^n)^T = x_k m \quad (22)$$

$m(r)$ is a scalar that $m(r) = \lambda_r D_{i_1 r}^1 \cdots D_{i_n r}^n$. And x_k is the system state vector and m is the prediction parameter, both of them are obtained by the tensor decomposition.

A.4 ANALYSIS OF SAMPLE COMPLEXITY FOR FORMULATING THE SYSTEM DYNAMIC TENSOR

The paper (Chen et al., 2020) does not analyze the sample complexity to construct \mathcal{D} , which is a $n+1$ multi-dimensional tensor $\mathcal{D} \in \mathbb{R}^{|\mathcal{T}_1| \times \dots \times |\mathcal{T}_n| \times |\mathcal{H}|}$. We give this analysis. The $\mathcal{D}_{i_1, \dots, i_n, k} := \Pr(t_{1i_1}, \dots, t_{ni_n} | \mathbf{h}_k)$ is an element of that tensor \mathcal{D} such that t_{1i_1} is the i_1 -th test of agent 1, similarly, the t_{ni_n} is the i_n -th test of agent n , and \mathbf{h}_k is the joint history.

Proposition 1. *In a n -agents system, assume every agent has the same observation and action space $|\mathcal{O}|, |\mathcal{A}|$, for a length- k test, to formulate a complete system dynamics tensor \mathcal{D} defined in equation 21. Assume each entry of the tensor needs S samples to give a sufficient estimation using Monte-Carlo roll-out method, then the total sample size is at least $(|\mathcal{O}||\mathcal{A}|)^{kn}S$, if the agents are homogeneous, in other words, they are permutation invariant such that identity does not matter, then the total sample size is $\binom{|\mathcal{O}||\mathcal{A}|^k+n-1}{n}S$.*

Proof. At one time step, for any agent, it has $|\mathcal{O}||\mathcal{A}|$ different combinations for the joint test, then for a length of k tests, it follows that $(|\mathcal{O}||\mathcal{A}|)^k$ number of different choices. Then the tensor \mathcal{D} would need $(|\mathcal{O}||\mathcal{A}|)^{kn}$ elements to cover all the possible length k tests. So the total sample size is $(|\mathcal{O}||\mathcal{A}|)^{kn}S$. If the agents are homogeneous, then the ordering does not matter, for each agent, we have $(|\mathcal{O}||\mathcal{A}|)^k$ number of different choices for length k test, so the total choices for n agents are $\binom{|\mathcal{O}||\mathcal{A}|^k+n-1}{n}$, then we need $\binom{|\mathcal{O}||\mathcal{A}|^k+n-1}{n}S$ samples. \square

The sample complexity is exponentially scaled with the number of agents and length of tests.

B ALGORITHM AND INTEGRATING WITH MARL METHOD

We first introduce some backgrounds on two multi-agent frameworks: Multi-agent Markov Decision Process (MMDP) and Multi-agent Partially observable Markov Decision Process (MPOMDP) since the algorithms in our paper are developed based on the framework of MPOMDP.

B.1 MULTI-AGENT MDP AND MULTI-AGENT POMDP MODEL

B.1.1 MMDP

MMDP model is a tuple $(S, N, \{O_i\}_{i \in [n]}, \{A_i\}_{i \in [n]}, T, R)$, where S and N are finite sets of states and agents. A_i is a finite set of actions available to agent i ; $T : S \times A_1 \times \dots \times A_n \times S \mapsto [0, 1]$ is a transition function; and $R : S \mapsto R$ is the reward function. Each agent i obtains reward as function of the state and agent's action $r_i : S \times A_i \mapsto R$, and receives a private observation from the state by the observation channel $o_i : S \mapsto O_i$. The state has distribution $d : S \mapsto [0, 1]$. Each agent aims to maximize its own total expected return $r_i = \sum_{t=0}^T \gamma^t r_i^t$ where γ is a discounted factor and T is the time horizon. In a shared reward situation, there is a team reward function $r : S \times A_1 \times \dots \times A_n \mapsto R$, agents aim to maximize one shared total expected return $r = \sum_{t=0}^T \gamma^t r^t$.

B.1.2 MPOMDP

A MPOMDP model is a tuple $(S, N, \{O_i\}_{i \in [n]}, \{A_i\}_{i \in [n]}, \{\Omega_i\}_{i \in [n]}, T, R)$, where $(S, A_i, T_i, O_i, \Omega_i, R)$ describe a single-agent POMDP. O_i is the set of observations the agent i can make. $\Omega_i : S \times A_i \times O_i \mapsto [0, 1]$ is the agent's observation channel function, which specifies probabilities of observations given agent's actions and resulting states. (S, A_i, T_i, R_i) describes a single agent MDP; and each agent i obtaines reward as function of the state and agen's action $r_i : S \times A_i \mapsto R$. Each agent aims to maximize its own total expected return $r_i = \sum_{t=0}^T \gamma^t r_i^t$ where γ is a discounted factor and T is the time horizon. In POMDP, an agent's belief about the sate is represented as probability distribution over S . The agent has prior belief $b_{0,i}$. The agent's current belief, $b_{t,i}$ over S , is continuously revised based on new observations and expected results of performed actions. The belief update takes into account changes in initial belief, $b_{t-1,i}$, due to action $a_{t,i}$, executed at time $t - 1$, and the new observation, $o_{t,i}$. The new time belief state can be obtained from basic probability theory as follows: $b_i(s_t) = \beta \Omega_i(o_{t,i}, s_t, a_{t-1,i}) \sum_{s_{t-1} \in S} b_{t-1,i}(s_{t-1}) T(s_t, a_{t,i}, s_{t-1})$, where β is the normalizing factor.

B.2 ALGORITHM: MAPSRL

We give the details of MAPSRL in Algorithm 1.

B.3 MAPSRL IN DIAGRAM

We also use a diagram (Fig 4) to depict the algorithm. The algorithm runs k iterations; each iteration first uses the policy to roll out data and uses the regressions to obtain the PSR parameters. Then it executes the policy phase by using the PSR as the input of policy to generate action and using the current PSR parameters to update the predictive state. At the end of the policy phase, it updates both the PSR and policy parameters. The policy parameters and the linear relationship captured by W_i and Z_i are parameterized by the neural network, section C.3 has the network architectures. The loss is a composite loss that includes loss from actor-critic and the loss from the predictive state representation. The next iteration will re-learn the PSR parameters using the newly generated data based on the current policy obtained from the previous iteration; then, it does a soft update to update the PSR parameters with ones obtained at the previous iteration.

B.4 INTEGRATING TWO COMMON MARL ALGORITHMS INTO MAPSR MODEL

Here we provide a brief intuitive introduction about how we connect existed MARL algorithms with MAPSR. Please look at Fig 4 for demonstration. $\psi_{t,i}^o, \psi_{t,j}^a$, and $\psi_{t,i,j}^h$ are embedding of $(o_{t:t+k-1,i}, a_{t:t+k-1,j})$, and $(o_{1:t-1,i}, a_{1:t-1,j})$. We also have embedding for extended part, labeled as ξ . $Q_{t,i,j}$ and $P_{t,i,j}$ are estimated by tensor regression using the embedding vectors. G represents the given

Algorithm 1 MAPSRL

```

1: Input: Learning rate  $\eta$ , a graph  $G$ , a static complete graph  $G_c$  or static non-compete graph  $G_s$  or dynamic graph  $G_t$ 
2: Initialize MARL Policy  $\Theta_{MARL}$  randomly
3: for  $k = 1, 2, 3, \dots$  iterations do
4:   State Tracking Phase
5:   Sample  $b = 1, 2, 3, \dots M$  batch of initial trajectories:  $\{(o_t^b, a_t^b)\}_{b=1}^M$  from existed policy obtained from previous iteration  $k - 1$ :  $\{\pi_i^{k-1}\}_{i=1}^n$ 
6:   Let  $\pi_{\theta_{MARL}} = \{\pi_i^{k-1}\}_{i=1}^n$  if available or the initial policy;
7:   Given  $G$ , calculate  $\hat{Q}_i$  and  $\hat{P}_i$ , and obtain the initial  $W_i, Z_i, F_i$ :
8:     (1).Regression  $Q_{t,i,j} = T_{i,j} \circ U_{i,j}(h_{t,i,j})$  to get the  $\hat{Q}_{t,i,j}$  and  $\hat{P}_{t,i,j}$ 
9:     (2).Given the graph  $G$ , using equation 1, 4, or 5 to obtain  $\hat{Q}_{t,i}, \hat{P}_{t,i}$ 
10:    (3).Given the  $\hat{Q}_{t,i}, \hat{P}_{t,i}$ , to obtain  $W_i, Z_i, F_i$ 
11:
12: Policy Phase
13: Initialize MAPSR parameters  $\Theta_{MAPSR} = \{Q_{1,i}, W_i, Z_i\}_{i=1}^n$  from state tracking phase and previous iteration by a soft-update:

$$W_i = \beta W_i + (1 - \beta) W_i^{k-1}, Z_i = \beta Z_i + (1 - \beta) Z_i^{k-1}, F_i = F_i$$

14: for  $b = 1, 2, 3, \dots M$  batch of trajectories from  $\{\pi_i^{k-1}\}_{i=1}^n$  do
15:   Reset episode:  $a_{0,i}^b, o_{0,i}^b$ 
16:   for  $t = 0, 1, 2, \dots T$  roll-in in each trajectory do
17:     for Each agent  $i$  do
18:       Get observation  $o_{t,i}^b$  and reward  $r_{t,i}^b$ 
19:       Extension  $P_{t,i}^b = W_i(Q_{t,i}^b)$ 
20:       Filtering  $Q_{t+1,i}^b = F_i(Q_{t,i}^b, a_{t,i}^b, o_{t,i}^b, W_i)$ 
21:       Execute  $a_{t+1,i}^b \sim \pi_i^{k-1}(Q_{t+1,i}^b)$ 
22:       Predict  $\hat{o}_{t,i}^b = Z_i(Q_{t,i}^b, a_{t,i}^b)$ 
23:       Collect  $o_{t,i}^b, \hat{o}_{t,i}^b, a_{t,i}^b, r_{t,i}^b, Q_{t,i}^b, Q_{t+1,i}^b$ 
24:     end for
25:   end for
26: end for
27: Update  $\Theta$  using  $D = \{\{o_{t,i}^b, \hat{o}_{t,i}^b, a_{t,i}^b, r_{t,i}^b, Q_{t,i}^b\}_{i=1}^n\}_{t=1}^T\}_{b=1}^M$ :

$$\Theta \leftarrow Update(\Theta^{k-1}, D, \eta)$$
 as in Equation (8), get  $W_i^k, Z_i^k$ , and  $\pi_i^k$ , the linear operator  $W_i, Z_i$ , and nonlinear operator  $\pi_i$  are all parameterized by neural network.
28: end for
29: Output: Return  $\Theta = (\Theta_{MAPSR}, \Theta_{MARL})$ 

```

graph, the estimation of $Q_{t,i}$ and $P_{t,i}$ are based on graph G and equation 1, equation 4 and equation 5 for reference. Please also go to subsection 3.2 and section 4.1.1 for a verbal description. The policy network uses the predictive state as input to return the action. The agent takes the action to get the observation. The filter F_i takes predictive state, action, and observation as inputs to get the next predictive state. We show a centralized critic and gives the description in paragraph B.4.2. The loss is composed into two parts, $\mathcal{L}_{\theta_{MAPSR}} = \|Z_i\{F_i[W_i(Q_{t-1,i}), a_{t-1,i}, o_{t-1,i}], a_{t,i}\} - o_{t,i}\|_2^2$, and $\mathcal{L}_{\theta_{critic}}$, which we give detailed explanation in equation 29. Please also go to section 5 and Algorithm B.2 for more details about the framework.

B.4.1 MAPSRL-1

It is based on the IAC (Foerster et al., 2018) which directly applies the single-agent policy gradient to have each agent learn independently, with the idea behind independent Q-learning (Tan, 1993), with actor-critic in place of Q-learning.

IAC trains an actor-critic pair for each agent, resulting in actors $\pi_i(a_i|o_i)$ and critics $V_i(o_i, a_i)$.

$$\nabla_{\theta_i} J(\pi) = \mathbb{E}_{\pi} \left[\nabla_{\theta_i} \log \pi_i(a_i|o_i) (r_i + \gamma V_i^{\pi_i}(o_{t+1,i}, a_{t+1,i}) - V_i^{\pi_i}(o_i, a_i)) \right] \quad (23)$$

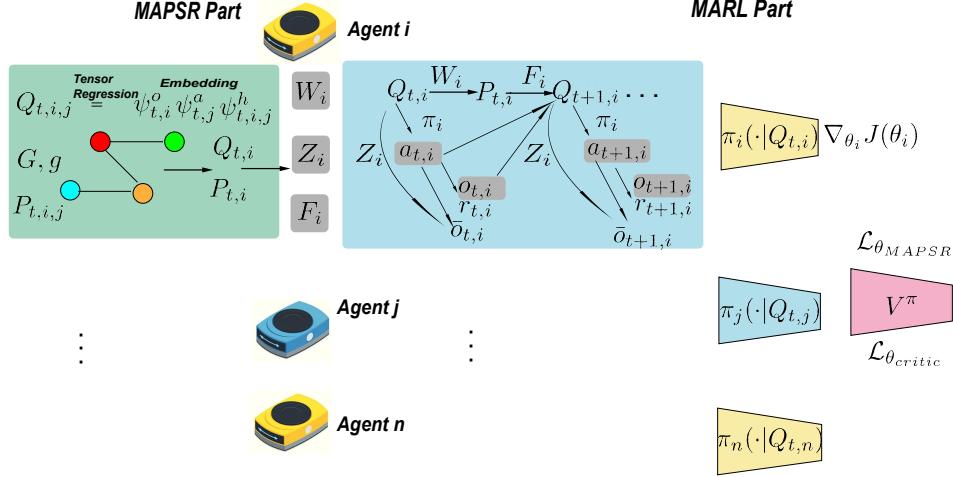


Figure 4: MAPSRL architecture combining MAPSR and MARL. The left part is the MAPSR model, which corresponds to the state tracking phase in Algorithm 1. The right part is the MARL, which corresponds to the policy phase. We use the actor critic framework, which contains a centralized critic and many decentralized actors. Both actor and critic are parameterized by neural networks. Section B.4 and section C.3 give a detailed verbal description about this architecture.

While IQL and IAC agents display a strong ability to optimize individual rewards (Tan, 1993), the lack of global information and a mechanism for cooperation means they are likely to settle for sub-optimal solutions.

Here we use the predictive state Q_i to fit the value and policy functions. And we train an actor-critic pair for each agent, resulting in actors $\pi_{i,\theta_i}(a_i|Q_i)$ and critics $V_i(Q_i)$.

$$\nabla_{\theta_i} J(\theta_i) = \mathbb{E}[\nabla_{\theta_i} \log \pi_{i,\theta_i}(a_i|Q_i)(r_i + \gamma V_i^{\pi_i}(Q_{t+1,i}) - V_i^{\pi_i}(Q_i))] \quad (24)$$

If the agents are homogeneous, we can share the critic and actor network. We have the critic loss as

$$\mathcal{L} = \mathbb{E}[(V^{\pi_i}(Q_i) - y_i)^2] \quad y_i = r_i + \gamma \hat{V}^{\pi_i}(Q_{t+1,i}) \quad (25)$$

Here \hat{V} is the target value function.

B.4.2 MAPSRL-2

The algorithm borrows the idea from MADDPG (Lowe et al., 2017) and LICA (Zhou et al., 2020) using the gradient of the PSR-value function with respect to policy to direct the policy gradient update.

MADDPG (Lowe et al., 2017) is an extension of deep deterministic actor-critic policy gradient (DDPG) (Lillicrap et al., 2015) to multi-agent setting such that let each agent's own critic is augmented with extra information about the actions of other agents, while their individual actor maintains a local state or observation. The gradient of each agent is:

$$\nabla_{\theta_i} J(\pi_i) = \mathbb{E}[\nabla_{\theta_i} \pi_{\theta_i}(a_i|o_i) \nabla_{\theta_i} V_i^a(o_1, \dots, o_n, a_1, \dots, a_n) | a_i = \pi(o_i)] \quad (26)$$

The action-value function V_i is updated as

$$\mathcal{L}_{\theta_i} = \mathbb{E}[(V_i^a(o_1, \dots, o_n, a_1, \dots, a_n) - y)^2] \quad y = r_i + \gamma \hat{V}_i^{a'}(o'_1, \dots, o'_n, a'_1, \dots, a'_n) \quad (27)$$

Where a' is the set of target policies with delayed parameters.

Unlike MADDPG using the deterministic policy, the gradient is taken on the policy parameter space, which is reflected in LICA Zhou et al. (2020).

$$\nabla_{\theta_i} J(\theta_i) = \mathbb{E}[\nabla_{\theta_i} \log \pi_{\theta_i}(a_i|Q_i) \nabla_{\theta_i} V^{\pi}(Q_1^{\pi_{\theta_1}}, \dots, Q_n^{\pi_{\theta_n}})] \quad (28)$$

Our method also uses the centralized critic to save space. Also, in order to solve the non-stationary environments when each agent is learning, it uses the joint predictive states as input without considering other agents' actions, which saves the input space since the predictive state is a conditional operator that considers other agents' actions.

The critic loss is defined as below:

$$\begin{aligned} \mathcal{L}_{\theta_{critic}} &= \mathbb{E}[(V^\pi(Q_1, \dots, Q_n) - y)^2] \\ y &= \sum_i r_i + \gamma \hat{V}^\pi(Q_{t+1,1}^{\pi_{\theta_1}}, \dots, Q_{t+1,n}^{\pi_{\theta_n}}) \end{aligned} \quad (29)$$

Even though the original MADDPG has the current actions as input to the value function so that the environment is stationary, we take out this operation because the predictive state already considers the effect actions can bring.

C ENVIRONMENT AND EXPERIMENT

C.1 MAMUJOCO ENVIRONMENT SETUP

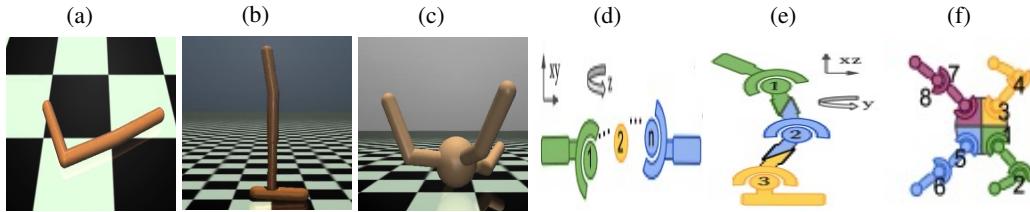


Figure 5: The illustration of three environments swimmer, hopper, ant, and their corresponding MAMuJoCo version. (a) Single swimmer; (b) Single hopper; (c) Single ant; (d) n-agents swimmer; (e) 3-agents hopper; (f) 4-agents ant.

Here, we give the details of setting up our multi-agent environment - MAMuJoCo. As introduced in section 6, the many agents are constructed by separating a existed single agent into parts, and each agent will only control a part of the whole agent (Figure 5).

Partially Observable space: MAMuJoCo is a simulated robotic environment, the partially observable property is achieved by only allowing partial information to the agents. For all environments, only the angles of the agent’s joints are visible to the network; the velocities are hidden.

Action: Each agent’s action space in MAMuJoCo is given by the joint action space overall motors controllable by that agent.

Observation: For each agent i , observations are constructed by inferring which body segments and joints are observable by an agent i . Each agent can always observe all joints within its sub-graph. A configurable parameter $k \geq 0$ determines the maximum graph distance to the agent’s subgraph at which joints are observable. For example, $k = 0$ means agents can only observe their own joints and body parts, while $k = 1$ means it can observe its adjacent joints, which has 1 graph distance to the agent. The agent observation is then given by a fixed order concatenation of each observable graph element’s representation vector. Depending on the environment and configuration, representation vectors may include attributes such as position, velocity, and external body forces. In addition to joint and body segment-specific observation categories, agents can also be configured to observe the robot’s central torso’s position and velocity attributes.

C.2 EXPERIMENT SETUP

We select three experiments from MAMuJoCo and give a detailed description of the experiments’ setup. In all environments, the agent has the goal to maximize the velocity of the first coordinate for the team. We use k to denote the maximum observation distances to the subgraph. We use Δd to denote the first coordinate position difference between a time difference Δt . Finally, we use R to denote the reward function. Table 1 has the configuration details for these parameters.

C.3 NEURAL NETWORK ARCHITECTURE

We implement all algorithms using deep neural networks as function approximators. We ensure that all policy, value, and action-value functions have the same neural network architecture among all algorithms to the extent each algorithm allows for a fair comparison.

Table 1: Configurations for MAMuJoCo environment

Environments		
Swimmer		
k	0	
R	$\sum_i (\frac{\Delta d_i}{\Delta t}) + 0.0001r$	
r	$r = -\ \mathbf{a}\ _2^2$ is a regularizer for joint action \mathbf{a}	
Hopper		
k	2	
R	$\sum_i (\frac{\Delta d_i}{\Delta t}) + 0.001r + 1.0$	
r	$r = -\ \mathbf{a}\ _2^2$ is a regularizer for joint action \mathbf{a}	
Ant		
k	0	
R	$\sum_i (\frac{\Delta d_i}{\Delta t}) + 5 \cdot 1e(-3) \ \text{external contact forces}\ _2^2 + 0.0001r$	
r	$r = -\ \mathbf{a}\ _2^2$ is a regularizer for joint action \mathbf{a}	

For our experiments with continuous action spaces, a Gaussian distribution with a diagonal covariance matrix is used. The policy network maps from the input feature to a Gaussian distribution vector μ . Moreover, $\mu = [\text{mean}, \text{std}]$, where mean is a vector specifies the action means, and std vector specifies the standard deviation.

In the implementation of all actor-critic method, all the actor-network is parameterized by a multi-layer perceptron (MLP) with two hidden layers of size 400 and 300 respectively and ReLU activation, which takes in the individual agent’s predictive state and outputs the mean and covariance of a Gaussian policy. The critic network is also an MLP with two hidden layers with 400 and 300 units, respectively. For MAPSRL-1, the critic network is used to approximate per-agent utilities, which receives each agent’s predictive state as input. For IAC, same as MAPSRL-1, it receives agent local observation and individual action as input.

In MAPSRL-2, there is a shared critic network that approximates all agents utilities, which receives all agents’ predictive states as input. In MADDPG, the critic receives the global state and the joint action of all agents as input. The global state consists of the complete state information from the original OpenAI Gym environment. Each decentralized actor (i.e., policy) network takes in each agent’s observation and outputs the agent’s action vector.

C.4 MODEL PARAMETERS

The hyper parameters for the the MAMuJoCo environments are in Table 2.

D SUPPLEMENT EXPERIMENTS

D.1 MAMUJOCO SUPPLEMENT EXPERIMENTAL RESULTS

We plotted the predictive observations compared to actual observations in Figure 2 at the beginning of the learning process (iteration 1) and end of the learning (iteration 40). We also show the results of iteration ten and iteration 20 in Figure 6. By comparing with iterations 1 and iterations 40 in Figure 2, we see that the iteration 40 has the smallest difference between predictive observation and true observation, and the difference gets increased as the iteration goes to the earlier stage of the learning process. So the predictive accuracy is improved incrementally with the learning progresses.

D.2 MULTI-AGENT PARTICLE ENVIRONMENT

We also test MAPSR into another environment, multi-agent particle environment (Lowe et al., 2017). The agents are displaced into a 2-dimensional coordinate. This environment does not assume that all agents have identical action and observation spaces. We run experiments using a different number of agents on two environments, the predator-prey, and cooperative-push. We use the same configuration

Table 2: Model parameters for MAMuJoCo environment

	Environments		
	Swimmer	Hopper	Ant
n	2	3	4
μ	0	0	0
σ	0.1	0.1	0.1
γ	0.99	0.99	0.99
Soft target network	0.001	0.001	0.001
α_1	0.7	0.65	0.7
α_2	0.3	0.35	0.3
β	0.6	0.45	0.5
η learning rate of Adam	0.001	0.001	0.001
λ ridge regression regularization	0.01	0.01	0.01
Total iterations	50	50	50
Number of trajectories	100	100	100
Maximum number of steps per trajectory	1000	1000	1000
Length of test window	8	12	10
Length of history window	8	12	10

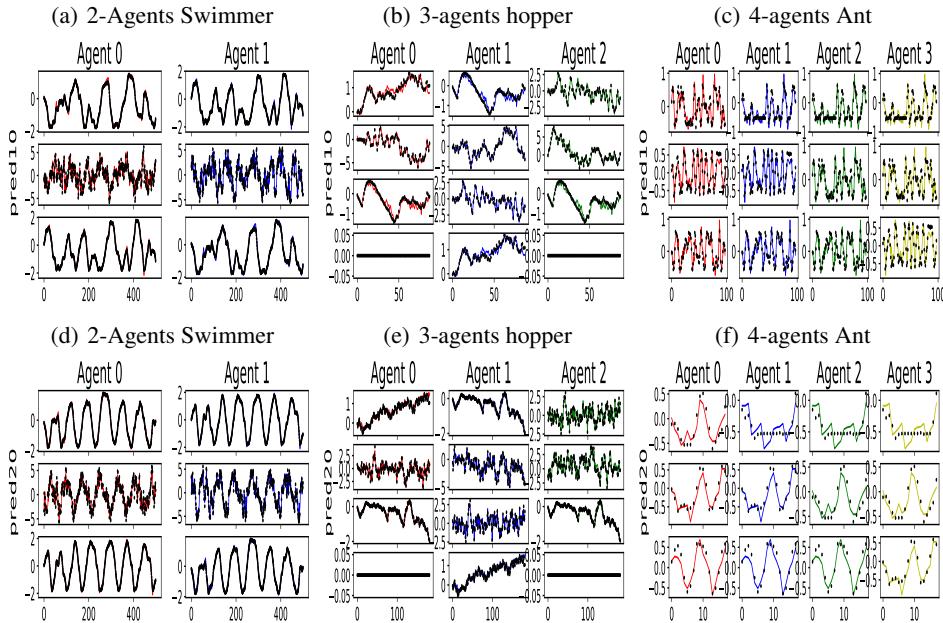


Figure 6: Additional Experiments Results, Predicted Trajectories vs Actual Observations for Multi-Agent Environments. (a) - (c) Iteration 10; (d) - (f) Iteration 20; This is the supplement for Figure 2

in (Liu et al., 2020a). The network has the same design as MAMuJoCo in section C.3, except we use MLP with two hidden layers with the same 128 units respectively, so we do not repeat the description. In Table 3, we report the hyperparameters for multi-agent particle environments.

E DETAILED PROOFS

We give the proof of Theorem 1 in Appendix E.1, and we first introduce the following Lemmas to prepare the proof.

Definition 2. Let $X_1 \dots X_k$ be independent random variables of dimensionality $d_{X_1} \dots d_{X_k}$ such that $\|X_k\| < t_{x_k}$. Let $\{(x_{1j}, \dots, x_{kj})\}_{j=1}^N$ be the N i.i.d samples from distribution of X_1, \dots, X_k ,

Table 3: Model parameters for multi-agent particle environment

	Environments					
	Predator-prey			Cooperative-push		
	n=3	n=15	n=100	n=3	n=15	n=30
γ	0.99	0.99	0.99	0.99	0.99	0.99
Soft target network	0.001	0.001	0.001	0.001	0.001	0.001
α_1	0.8	0.75	0.5	0.8	0.75	0.7
α_2	0.2	0.25	0.5	0.2	0.25	0.3
β	0.45	0.5	0.45	0.55	0.5	0.5
η learning rate of Adam	0.05	0.05	0.05	0.05	0.05	0.05
λ ridge regression regularization	0.01	0.01	0.01	0.01	0.01	0.01
Total iterations	30	30	30	30	30	30
Number of trajectories	100	100	100	100	100	100
Maximum number of steps per trajectory	500	500	500	500	500	500
Length of test window	5	5	5	5	5	5
Length of history window	5	5	5	5	5	5

the $C_{X_1} := \mathbb{E}[X_1 X_1^T]$ and $\widehat{C}_{X_1} = \frac{1}{N} \sum_{j=1}^N x_{1j} x_{1j}^T$, and $C_{X_1 X_2} := \mathbb{E}[X_1 X_2^T]$ and $\widehat{C}_{X_1, X_2} = \frac{1}{N} \sum_{j=1}^N x_{1j} x_{2j}^T$. Also, we use the $\cdot, v(\cdot)$ to denote the largest, smallest eigenvalue of a matrix.

Lemma 3 ((Tropp, 2015)). Let a_j be a finite sequence of independent random, Hermitian matrices with dimension d . Assume that $0 \leq v(a_j)$ and $u(a_j) \leq L$ for each j . Let $S = \sum_j a_j$, then for any $\eta \in [0, 1]$, it follows that

$$\Pr(v(S) \leq (1 - \eta)v(\mathbb{E}[S])) \leq d \left[\frac{e^{-\eta}}{(1 - \eta)^{1-\eta}} \right]^{v(\mathbb{E}[S])/L} \leq 2d e^{-\eta v(\mathbb{E}[S])/L} \quad (30)$$

Corollary 1. Let X be a random variable, for any $\epsilon \in (0, 1)$ such that $N > \frac{t_x^2 \log(2d_X/\epsilon)}{v(C_X)}$ the following holds with probability at least $1 - \epsilon$

$$v(\widehat{C}_X) > \frac{t_x^2 \log(2d_X/\epsilon)}{v(C_X)N}$$

In other words, if N large enough, then \widehat{C}_X and C_X will be close enough.

Proof. Define $S_j = 1/N x_j x_j^T$. Then it follows that $u(S_j) \leq L = t_x^2/N$ and define $\epsilon := 2d_X e^{-\sigma N v(C_X)/t_x^2}$, which implies that $\sigma = \frac{t_x^2 \log(2d_X/\epsilon)}{v(C_X)N}$. Then it follows from Matrix Chernoff Inequality in Lemma 3 that $P_r(v(\widehat{C}_X) \leq (1 - \sigma)v(C_X)) \leq \epsilon$ \square

Lemma 4 ((Tropp, 2015)). A finite sequence $\{a_j\}$ of independent, random matrices with common dimensions $a \times b$, and assume that $\mathbb{E}[a_j] = 0$ and $\|a_j\| \leq L$ for each j , let $S = \sum_j a_j$ as a random matrix. Let $\text{Var}(S)$ be the variance statistics such that $\text{Var}(S) = \max\{\|\mathbb{E}[SS^T]\|, \|\mathbb{E}[S^T S]\|\}$, then

$$\Pr(\|S\| > c) \leq (a + b) e^{-\frac{c^2/2}{\text{Var}(S) + Lc/3}} \quad (31)$$

Corollary 2. With at least probability $1 - \epsilon$ that

$$\|\widehat{C}_{YX} - C_{YX}\| \leq \sqrt{\frac{2 \log(d_Y + d_X)/\epsilon \text{Var}}{N}} + \frac{2 \log((d_Y + d_X)/\epsilon)L}{3N}$$

where $L = t_y t_x + \|C_{YX}\| \leq 2t_y t_x$ and $\text{Var} = \max\{t_y^2 \|C_X\|, t_x^2 \|C_Y\|\} + \|C_{YX}\|^2 \leq 2t_y^2 t_x^2$

Proof. Let X, Y be two random variables, and let a finite sequence $\{a_j\}$ of independent random matrices to satisfy $a_j = y_j x_j^T - C_{YX}$. So the a_j will have dimensions $d_X \times d_Y$. Let random matrix $S = \sum_j a_j$. It follows that $\mathbb{E}[a_j] = 0$ and $\|a_j\| = \|y_j x_j^T - C_{YX}\| \leq \|y_j\| \|x_j\| + \|C_{YX}\| \leq$

$$\begin{aligned}
& t_y t_x + \|C_{YX}\| \\
\|\mathbb{E}[SS^T]\| &= \left\| \sum_{i,j} (\mathbb{E}[y_i x_i^T x_j y_j^T] - C_{YX} C_{XY}) \right\| \\
&= \left\| \sum_i (\mathbb{E}[\|x_i\|^2 y_i y_i^T] - C_{YX} C_{XY}) + \sum_{i \neq j} (\mathbb{E}[y_i x_i^T] \mathbb{E}[x_j y_j^T] - C_{YX} C_{XY}) \right\| \\
&= \left\| \sum_i (\mathbb{E}[\|x_i\|^2 y_i y_i^T] - C_{YX} C_{XY}) \right\| \\
&\leq N(t_x^2 \|C_Y\| + \|C_{YX}\|^2)
\end{aligned}$$

Similarly, $\|\mathbb{E}[SS^T]\| \leq N(t_y^2 \|C_X\| + \|C_{YX}\|^2)$. By applying lemma 4, we have $\epsilon = \mathbf{Pr}(\|S\| \geq Nc) \leq (d_X + d_Y) e^{\frac{-Nc^2/2}{Var+Lc/3}}$ and therefore, it implies that

$$\begin{aligned}
c &\leq \frac{\log((d_X + d_Y)/\epsilon)L}{3N} + \sqrt{\frac{(\log(d_X + d_Y)/\epsilon))^2 L^2}{9N^2} + \frac{2\log((d_X + d_Y)/\epsilon)Var}{N}} \\
&\leq \frac{2\log((d_X + d_Y)/\epsilon)L}{3N} + \sqrt{\frac{2\log((d_X + d_Y)/\epsilon)Var}{N}}
\end{aligned}$$

□

Corollary 3. For random variable X with dimensionality d_X and $\|X\| \leq t_x$, with probability $1 - \epsilon$, it follows that

$$\|C_X^{-1/2}(\hat{C}_X) - C_X\| \leq 2t_x \sqrt{\frac{2\log(2d_X/\epsilon)}{N} + \frac{2\log(2d_X/\epsilon)L}{3N}}$$

where $L = \frac{t_x^2}{\sqrt{v(C_X)}} + t_x$

Proof. The proof is similarly to the the proof of corollary 2, define $a_j = \sum_X^{-1/2} x_j x_j^T - C_X^{1/2}$, $S = \sum a_j$ then it follows that $\mathbb{E}[a_j] = 0$ and $\|a_j\| \leq \frac{t_x^2}{\sqrt{v(C_X)}} + t_x$

$$\|\mathbb{E}[S^T S]\| = \|\mathbb{E}[SS^T]\| \leq N(t_x^2 + \|C_X\|^2) \leq 2Nt_x^2 \quad (32)$$

Applying lemma 4 to get

$$\epsilon = \mathbf{Pr}(\|S\| \geq Nc) \leq 2d_X e^{\frac{-Nc^2/2}{2t_x^2 + Lc/3}} \quad (33)$$

it follows that

$$c \leq \frac{2\log(2d_X/\epsilon)L}{3N} + 2t_x \sqrt{\frac{\log(2d_X/\epsilon)}{N}} \quad (34)$$

□

Lemma 5. For two random variables X, Y , let $\hat{C}_{YX} = C_{YX} + \Delta_{YX}$, and $\hat{C}_X = C_X + \Delta_X$ where $\mathbb{E}[\Delta_{YX}]$ and $\mathbb{E}[\Delta_X]$ are not necessarily zero and \hat{C}_X is symmetric positive semidefinite. Define $A = C_{YX} C_X^{-1}$ and $\hat{A} = \hat{C}_{YX} (\hat{C}_X + \lambda)^{-1}$. Then it follows that:

$$\|\hat{A} - A\| \leq \sqrt{\frac{u(C_Y)}{v(C_X)}} \left(\frac{\sqrt{v(C_X)} \|C_X^{-1/2}\| \Delta_X + \lambda}{v(\hat{C}_X) + \lambda} \right) + \frac{\|\Delta_{YX}\|}{v(\hat{C}_X) + \lambda}$$

Proof.

$$\hat{A} - A = C_{YX} ((C_X + \Delta_X + \lambda I)^{-1} - C_X^{-1}) + \Delta_{YX} (C_X + \Delta_X + \lambda I)^{-1} = M_1 + M_2$$

It follows that

$$\|M_2\| \leq \frac{\Delta_{YX}}{v(\hat{C}_X) + \lambda}$$

For M_1 , by using facts $U^{-1} - V^{-1} = U^{-1}(V - U)V^{-1}$ and $C_{YX} = C_Y^{1/2}PC_X^{1/2}$, where P is a correlation matrix with $\|P\| \leq 1$,

$$\begin{aligned} M_1 &= -C_{YX}C_X^{-1}(\Delta_X + \lambda I)(C_X + \Delta_X + \lambda I)^{-1} \\ &= -C_Y^{1/2}PC_X^{-1/2}(\Delta_X + \lambda I)(C_X + \Delta_X + \lambda I)^{-1} \end{aligned}$$

$$\begin{aligned} \|M_1\| &\leq \sqrt{u(C_Y)} \frac{\|C_X^{-1/2}\Delta_X\| + \lambda \|C_X^{-1/2}\|}{v(\hat{C}_X) + \lambda} \\ &= \sqrt{\frac{u(C_Y)}{v(C_X)}} \frac{\sqrt{v(C_X)} \|C_X^{-1/2}\Delta_X\| + \lambda}{v(\hat{C}_X) + \lambda} \end{aligned}$$

□

Corollary 4. Let $\{(x_k, y_k)\}_{k=1}^N$ be i.i.d samples from two random variables X, Y with dimensions d_X and d_Y and (uncentered) covariances C_X and C_Y . Assume $\|X\| \leq t_x$ and $\|Y\| \leq t_y$. Define $A = C_{YX}C_X^{-1}$ and $\hat{A} = \hat{C}_{YX}(\hat{C}_X + \lambda)^{-1}$. For any $\epsilon \in (0, 1)$ such that $N > \frac{t_x^2 \log(2d_X/\epsilon)}{v(C_X)}$ the following holds with probability at least $1 - 3\epsilon$

$$\|\hat{A} - A\| \leq \sqrt{\frac{u(C_Y)}{v(C_X)}} \left(\frac{\sqrt{v(C_X)}\alpha + \lambda}{v(C_X)(1 - \gamma) + \lambda} \right) + \frac{\beta}{v(C_X)(1 - \gamma) + \lambda}$$

where

$$\alpha = 2t_x \sqrt{\frac{2 \log(2d_X/\epsilon)}{N}} + \frac{2 \log(2d_X/\epsilon)}{3N} \left(\frac{c_x^2}{\sqrt{v(C_X)}} + t_x \right)$$

$$\beta = 2t_y t_x \sqrt{\frac{\log(d_Y + d_X)/\epsilon}{N}} + \frac{3t_y t_x \log((d_Y + d_X)/\epsilon)}{3N}$$

$$\gamma = \frac{t_x^2 \log(2d_X/\epsilon)}{v(C_X)N}$$

Proof. It follows by applying Corollaries 1, 2, 3 to Lemma 5. By union bound, each condition has probability $1 - \epsilon$, so the total events are bounded

$$\Pr(\text{bounds satisfied}) := 1 - \Pr\left(\bigcup_{i=1}^3 A_i\right) \geq 1 - \sum_{i=1}^3 \Pr(A_i) = 1 - 3\epsilon$$

□

Lemma 6. For two random variables X, Y , let $\hat{C}_{YX} = C_{YX} + \Delta_{YX}$, and $\hat{C}_X = C_X + \Delta_X$ where $\mathbb{E}[\Delta_{YX}]$ and $\mathbb{E}[\Delta_X]$ are not necessarily zero and \hat{C}_X is symmetric but not positive semidefinite. Define $A = C_{YX}C_X^{-1}$ and $\hat{A} = \hat{C}_{YX}\hat{C}_X(\hat{C}_X^2 + \lambda)^{-1}$. Then it follows that:

$$\|\hat{A} - A\| \leq \sqrt{\frac{u(C_Y)}{v(C_X)^3} \frac{\|\Delta_x\|^2 + 2u(C_X)\|\Delta_X\| + \lambda}{v(\hat{C}_X) + \lambda}} + \frac{\|C_{YX}\|\|\Delta_X\| + \|\Delta_{YX}\|\|C_X\| + \|\Delta_{YX}\|\|\Delta_X\|}{v(\hat{C}_X)^2 + \lambda}$$

Proof.

$$\begin{aligned} \hat{A} - A &= (C_{YX} + \Delta_{yx})(C_X + \Delta_X)((C_X + \Delta_X)^2 + \lambda I)^{-1} - C_{YX}C_XC_X^{-2} \\ &= C_{YX}C_X(((C_X + \Delta_X)^2 + \lambda I)^{-1} - C_X^{-2}) + (C_{YX}\Delta_X + \Delta_{YX}C_X + \Delta_{YX}\Delta_X)((C_X + \Delta_X)^2 + \lambda I) \\ &= M_1 + M_2 \end{aligned}$$

For M_1 , by using facts $U^{-1} - V^{-1} = U^{-1}(V - U)V^{-1}$ and $C_{YX} = C_Y^{1/2}PC_X^{1/2}$, where P is a correlation matrix with $\|P\| \leq 1$, it follows that

$$M_1 = -C_Y^{1/2}PC_X^{-3/2}(\Delta_X^2 + C_X\Delta_X + \Delta_XC_X + \lambda I)((C_X + \Delta_X)^2 + \lambda I)^{-1}$$

Therefore,

$$\|M_1\| \leq \sqrt{\frac{u(C_Y)}{v(C_X)^3} \frac{\|\Delta_x\|^2 + 2u(C_X)\|\Delta_x\| + \lambda}{v(\widehat{C}_X) + \lambda}}$$

$$\|M_2\| \leq \frac{\|C_{YX}\|\|\Delta_x\| + \|C_{YX}\|\|C_X\| + \|C_{YX}\|\|\Delta_x\|}{v(\widehat{C}_X)^2 + \lambda}$$

□

Corollary 5. Let $\{(x_k, y_k)\}_{k=1}^N$ be i.i.d samples from two random variables X, Y with dimensions d_X and d_Y and (uncentered) covariances C_X and C_Y . The $\mathbb{E}[\Delta_{YX}]$ and $\mathbb{E}[\Delta_X]$ is not necessarily zero and C_x is symmetric but not necessarily positive semidefinite. Assume $\|X\| \leq t_x$ and $\|Y\| \leq t_y$. Define $A = C_{YX}C_X^{-1}$ and $\widehat{A} = \widehat{C}_{YX}\widehat{C}_X(\widehat{C}_X^2 + \lambda)^{-1}$. For any $\epsilon \in (0, 1)$ such that $N > \frac{t_x^2 \log(2d_X/\epsilon)}{v(C_X)}$ the following holds with probability at least $1 - 3\epsilon$

$$\|\widehat{A} - A\| \leq \sqrt{\frac{u(C_Y)}{v(C_X)^3} \frac{\|\Delta_x\|^2 + 2u(C_X)\|\Delta_x\| + \lambda}{v(C_X)(1 - \gamma) + \lambda}} + \frac{\|C_{YX}\|\|\Delta_x\| + \|\Delta_{YX}\|\|C_X\| + \|\Delta_{YX}\|\|\Delta_x\|}{v(C_X)^2(1 - \gamma)^2 + \lambda}$$

where

$$\gamma = \frac{t_x^2 \log(2d_X/\epsilon)}{v(C_X)N}$$

Proof. It follows by applying Corollaries 1,2,3 to Lemma 6. Also by union bound, so the total events are bounded $1 - 3\epsilon$

□

Theorem 2. Grabner 1997 (3.3) (Grabner & Prodinger, 1997) Consider there are n independent copies X_1, \dots, X_n i.i.d negative binomial random variables, with parameters defined as $\mathcal{NB}(b, p)$, and our goal is to calculate the expectation of the maximum of these N random variables $\mathbb{E}_n = \mathbb{E}\{\max(X_1, \dots, X_n)\}$ then we have following asymptotic solution

$$\begin{aligned} \mathbb{E}_n = & \log_{\frac{1}{q}}(n) + (b-1) \log_{\frac{1}{q}} \log_{\frac{1}{q}}(n) + (b-1) \log_{\frac{1}{q}} p + (b-1) - \log_{\frac{1}{q}}(b-1)! + \frac{1}{2} + \frac{\gamma}{\log_{\frac{1}{q}}(1/q)} \\ & + F(\log_{\frac{1}{q}}(n) + (b-1) \log_{\frac{1}{q}} \log_{\frac{1}{q}}(n) + (b-1) \log_{\frac{1}{q}} p - \log_{\frac{1}{q}}(b-1)!) + o(1) \end{aligned} \quad (35)$$

(where F is a periodic C^∞ -function of period 1 and mean value 0 whose Fourier-coefficients are given by $\hat{F}(k) = -\frac{1}{\log(\frac{1}{q})}\Gamma(-\frac{2k\pi i}{\log(\frac{1}{q})})$ for $k \in \mathbb{Z} \setminus \{0\}$, and $q = 1 - p$)

We omit the proof, interested readers could go to (Grabner & Prodinger, 1997) for details.

E.1 PROOF OF THEOREM 1

We first prove the bound for $\|\widehat{Q}_{i,j} - Q_{i,j}\|$.

Proposition 2. Let π_Θ be a data collection policy and \mathcal{H} is the range of π_Θ on joint histories. If Equation 2 used, then for all $h \in \mathcal{H}$ and any $\epsilon \in (0, 1)$, $\|\widehat{Q}_{i,j}(\psi^h) - Q_{i,j}(\psi^h)\|$ is bounded as below with probability at least $1 - 3\epsilon$.

$$\begin{aligned} \|\widehat{Q}_{i,j} - Q_{i,j}\| \leq & \sqrt{\frac{u(C_{\psi_i^o | \psi_{i,j}^h})}{v(C_{\psi_j^a | \psi_{i,j}^h})^3} \frac{\|\Delta_1\|^2 + 2u(C_{\psi_j^a | \psi_{i,j}^h})\|\Delta_1\| + \lambda}{v(C_{\psi_j^a | \psi_{i,j}^h})(1 - \gamma) + \lambda}} \\ & + \frac{\|C_{\psi_i^o \psi_j^a | \psi_{i,j}^h}\| \|\Delta_1\| + \|\Delta_2\| \|C_{\psi_j^a | \psi_{i,j}^h}\| + \|\Delta_2\| \|\Delta_1\|}{v(C_{\psi_j^a | \psi_{i,j}^h})^2 (1 - \gamma)^2 + \lambda} \end{aligned}$$

where Δ_1 follows the bound 38 and Δ_2 follows the bound 36, and $\gamma = \frac{t_{A_j}^2 \log(2d_{A_j}/\epsilon)}{v(C_{\psi_j^a})N}$

Proof. Let $T_{i,j}$ is the tensor such that $C_{\psi_i^o \psi_j^a | \psi_{i,j}^h} = T_{i,j} \times_h \psi_{i,j}^h$, and $U_{i,j}$ is the tensor such that $C_{\psi_j^a | \psi_{i,j}^h} = U_{i,j} \times_h \psi_{i,j}^h$ and for simplicity without loss meaning, we use $C_{\psi_j^a | \psi_{i,j}^h}$ to denote $C_{\psi_j^a \psi_j^a | \psi_{i,j}^h}$. Then we have

$$\begin{aligned}\|\widehat{C}_{\psi_i^o \psi_j^a | \psi_{i,j}^h} - C_{\psi_i^o \psi_j^a | \psi_{i,j}^h}\| &\leq \|\widehat{T}_{i,j} - T_{i,j}\| \|\psi_{i,j}^h\| \\ \|\widehat{C}_{\psi_j^a | \psi_{i,j}^h} - C_{\psi_j^a | \psi_{i,j}^h}\| &\leq \|\widehat{U}_{i,j} - U_{i,j}\| \|\psi_{i,j}^h\|\end{aligned}$$

We finish the above proof by proofing the $\|T_{i,j} - \widehat{T}_{i,j}\|$ and $\|U_{i,j} - \widehat{U}_{i,j}\|$ are bounded by using Corollary 4.

$$\|\widehat{T}_{i,j} - T_{i,j}\| \|\psi_{i,j}^h\| \leq t_h \sqrt{\frac{u(C_{\psi_i^o, \psi_j^a})}{v(C_{\psi_{i,j}^h})}} \left(\frac{\sqrt{v(C_{\psi_{i,j}^h})} \alpha + \lambda}{v(C_{\psi_{i,j}^h}) (1 - \gamma) + \lambda} \right) + \frac{\beta}{v(C_{\psi_{i,j}^h}) (1 - \gamma) + \lambda}, \quad (36)$$

where

$$\begin{aligned}\alpha &= 2t_h \sqrt{\frac{2 \log(2d_h/\epsilon)}{N}} + \frac{2 \log(2d_h/\epsilon)}{3N} \left(\frac{t_h^2}{\sqrt{v(C_{\psi_{i,j}^h})}} + t_h \right), \\ \beta &= 2t_{O_i} t_{A_j} t_h \sqrt{\frac{\log(d_{O_i} d_{A_j} + d_h)/\epsilon}{N}} + \frac{4t_{O_i} t_{A_j} t_h \log((d_{O_i} d_{A_j} + d_h)/\epsilon)}{3N}, \\ \gamma &= \frac{t_h^2 \log(2d_h/\epsilon)}{v(C_{\psi_{i,j}^h}) N};\end{aligned} \quad (37)$$

and

$$\|\widehat{U}_{i,j} - U_{i,j}\| \|\psi_{i,j}^h\| \leq t_h \sqrt{\frac{u(C_{\psi_j^a})}{v(C_{\psi_{i,j}^h})}} \left(\frac{\sqrt{v(C_{\psi_{i,j}^h})} \alpha + \lambda}{v(C_{\psi_{i,j}^h}) (1 - \gamma) + \lambda} \right) + \frac{\beta}{v(C_{\psi_{i,j}^h}) (1 - \gamma) + \lambda}, \quad (38)$$

where

$$\begin{aligned}\alpha &= 2t_h \sqrt{\frac{2 \log(2d_h/\epsilon)}{N}} + \frac{2 \log(2d_h/\epsilon)}{3N} \left(\frac{t_h^2}{\sqrt{v(C_{\psi_{i,j}^h})}} + t_h \right), \\ \beta &= 2t_{A_j} t_h \sqrt{\frac{\log(d_{A_j} + d_h)/\epsilon}{N}} + \frac{4t_{A_j} t_h \log((d_{A_j} + d_h)/\epsilon)}{3N}, \\ \gamma &= \frac{t_h^2 \log(2d_h/\epsilon)}{v(C_{\psi_{i,j}^h}) N}.\end{aligned} \quad (39)$$

Then using the equation 2 and corollary 5 to obtain the bound for $Q_{i,j}$

$$\begin{aligned} \|\hat{Q}_{i,j} - Q_{i,j}\| &\leq \sqrt{\frac{u(C_{\psi_j^o|\psi_{i,j}^h})}{v(C_{\psi_j^a|\psi_{i,j}^h})^3} \|\Delta_1\|^2 + 2u(C_{\psi_j^a|\psi_{i,j}^h}) \|\Delta_1\| + \lambda} \\ &+ \frac{\|C_{\psi_i^o\psi_j^a|\psi_{i,j}^h}\| \|\Delta_1\| + \|\Delta_2\| \|C_{\psi_j^a|\psi_{i,j}^h}\| + \|\Delta_2\| \|\Delta_1\|}{v(C_{\psi_j^a|\psi_{i,j}^h})^2 (1-\gamma)^2 + \lambda}, \end{aligned}$$

where Δ_1 follows the bound 38 and Δ_2 follows the bound 36, and $\gamma = \frac{t_{A_j}^2 \log(2d_{A_j}/\epsilon)}{v(C_{\psi_j^a})^N}$

□

Now we start to prove our theorem 1.

Proof. The equation 1 says:

$$Q_{t,i} := g(\{Q_{t,i,j}\}_{j=1}^n) = \sum_j Q_{t,i,j},$$

here we assume the agents are homogeneous, in other words, each pair of $\{(o_i, a_i)\}_{i=1}^n$ coming from the same spaces \mathcal{O}, \mathcal{A} . They are permutation invariant and their identities do not matter. Thus, the bound of $Q_{i,j}$ is invariant to agents. Under the assumption of static fully complete graph, for any agent $Q_i = \sum_{j=1}^n Q_{i,j}$, thus

$$\begin{aligned} \|\hat{Q}_i - Q_i\| &= \left\| \sum_j \hat{Q}_{i,j} - \sum_j Q_{i,j} \right\| \\ &\leq n \|\hat{Q}_{i,j} - Q_{i,j}\| \end{aligned} \tag{40}$$

□

E.2 PROOF OF LEMMA 1

Here we prove Lemma 1.

Proof. Equation 4 says

$$Q_{t,i} := g(\{Q_{t,i,j}\}_{j=1}^n) = \sum_j I_{i,j} Q_{t,i,j},$$

where $I_{i,j}$ is an indicator function to denote if two agents are connected. Under the assumption that the static non-complete graph has maximum of number of degrees k and the agents are homogeneous, we have $\sum_j I_{i,j} \leq k$. Thus,

$$\begin{aligned} \|\hat{Q}_i - Q_i\| &= \left\| \sum_j I_{i,j} \hat{Q}_{i,j} - \sum_j I_{i,j} Q_{i,j} \right\| \\ &\leq k \|\hat{Q}_{i,j} - Q_{i,j}\| \end{aligned} \tag{41}$$

□

E.3 INFORMAL PROOF OF LEMMA 2

Here we make some intuitions for the proof of Lemma 2. As we already give the proof sketch in our main paper. Lemma 2 is a direct application of Theorem 2. Here our random variable $J_1, \dots, J_n \sim \mathcal{NB}(r, p)$, where $\{(J_i)\}_{i=1}^n$ represents the number of time points node i needs before it meets node j number of r times.

For the complete static graph, we need at least N sample for the bound in equation 3 to be valid; in other words, we need the trajectory to run at least N time points to collect enough data to estimate our conditional operator accurately $Q_{i,j}$.

For the dynamic graph, each time t , the two nodes are randomly connected with probability p , if it connects, then we can obtain a valid sample to estimate $Q_{i,j}$; if not, then we skip to the next

time step. For the dynamic graph node i , if we take the union set of the nodes i connected over the trajectory path, then the union set could form a static complete graph. We say the two graphs are equivalent. The number of time points needed by i until the N^{th} connection with j for each of the pairs (i, j) follows the same distribution $J \sim \mathcal{NB}(N, p)$. Then we are interested in the expectation of the maximum of J_1, \dots, J_{n-1} , we denote it as $J_{\{1, \dots, n-1\}}$.

$\mathbb{E}\{J_{\{1, \dots, n-1\}}\}$ means on average, how many time points (N') we need for all nodes other than i at least meets N times with i . Obviously, this $N' \geq N$ since $p \in [0, 1]$. The calculation of this expectation is solved by (Grabner & Prodinger, 1997). And we also put their result in Theorem 2.

Classification of Phonocardiogram Signals Based on Envelope Extraction

Lijun Yang^{a,*}, Shuang Li^a, Zhi Zhang^b, Xiaohui Yang^a,

^aSchool of Mathematics and Statistics, Henan University, Kaifeng 475004, China.

^bDepartment of Computer Science, Georgia Institute of Technology, Atlanta, GA, 30332, USA.

Abstract

The prevention and diagnosis of cardiovascular diseases have become one of the primary problems in the medical community since this kind of diseases mortality accounts for 31% of all global deaths in 2016. Heart sound is one of the most important physiological signal in human body, in which it contains lots of pathological information about each part of human heart. Therefore, heart sound signal analysis is an effective and convenient method for the preliminary diagnosis of heart disease. Phonocardiogram (PCG) is the representation of heart sounds in the form of graphs recorded by phonocardiograph. PCG signals have attracted attentions because that they comprise of bio-acoustic statistics indicating the functional condition of the heart. In this paper, we focus on detecting the normal and abnormal PCG signals. After using empirical mode decomposition (EMD) process to each PCG signal, the second intrinsic mode function is chosen for further analysis. Based on the proposed envelope model, the cardiac cycles of PCG signals can be estimated and then the time-domain features can be extracted. Combining with the frequency-domain features and wavelet-domain features, the feature vectors are obtained. Finally, SVM classifier is used to classify the normal and abnormal PCG signals. The proposed framework is evaluated on the public datasets and reaches the classification accuracies of more than 96%, which show the effectiveness of the proposed method.

Keywords: Heart Sound, Phonocardiogram (PCG), Empirical Mode Decomposition (EMD), Envelope Extraction, PCG Segmentation, Support Vector Machine

1. Introduction

According to World Health Organization, about 17.9 million people died from cardiovascular diseases (CVDs) in 2016, which represents 31% of all global deaths [20]. In China, CVD prevalence is on the rise and its mortality ranks first, higher than that of tumors and other diseases [8]. Therefore, the prevention and diagnosis of CVDs have become one of the primary problems in the medical community. The heart sound is one of the most important physiological signal in human body, in which it contains lots of pathological information about each part of human heart, such as atrial and ventricular, large blood vessels, cardiovascular and valvular function state [14]. Many pathological conditions of the cardiovascular system are reflected in the heart sound. Therefore, it is possible to diagnose heart diseases through heart sound analysis. Heart sound can be diagnosed by heart auscultation, which is a non-invasive, low-cost screening method and easily accepted by patients. However, heart sound analysis by auscultation depends highly on the skills and experience of the listener. Therefore, a computer assisted system can help the general physician in coming up to a more accurate and reliable diagnosis at early stages [39].

Phonocardiogram (PCG) is the representation of heart sounds in the form of graphs recorded by phonocardiograph [24]. PCG signals have attracted attentions because that they comprise of bio-acoustic statistics indicating the functional condition of the heart. Intelligent and automated analysis of the PCG is therefore very important not only in detection of cardiac diseases but also in monitoring the effect of certain cardiac drugs on the condition of the heart. PCG permits the examination of features to wider extent than auscultation. PCG is comparatively new metric and correct localization and classification of heart sounds has been a challenging task because of inconsistency of heart cycles. Many researchers have been trying to analyze PCG signals using different methods like filtering, transform based on algorithm, de-noising, feature extraction and classification. Major aims of research on PCG signals are to localize heart sounds and then to classify signals as normal or abnormal.

*Corresponding author

Email addresses: yanglijun@henu.edu.cn (Lijun Yang), lswwrain@163.com (Shuang Li), zhizhang@gatech.edu (Zhi Zhang), xhyang@henu.edu.cn (Xiaohui Yang)

In the last decade, lots of research work on PCG signal analysis has been carried out, which is mainly divided
25 into the following aspects:

- Noise removal

Heart sound signals are always disturbed by various factors and noise during acquisition process, including lung sounds, muscle contraction, breath sounds, and background noise. All these noise components can make subsequent PCG signal analysis difficult. Therefore, it is important to use appropriate denoising algorithm on PCG signals prior to processing. The essence of denoising is to find a suitable filter to separate the PCG signal from the noise. In reference [13], the authors used the least mean square algorithm to develop an adaptive filter for noise removal. In [26], the denoised signals were obtained by using median and low pass filter (Butterworth) on normalized PCG signals. Some researchers considered the decomposition methods for noise removal. The authors in [1, 22] applied wavelet transform on PCG denoising based on the fact that their wavelet coefficients tend to be much larger than those due to noise. Empirical Mode Decomposition (EMD) method was used to PCG signal denoising in [30]. After EMD processing, the PCG signal can be decomposed into several intrinsic mode functions (IMFs). Then, the energy density of the IMFs were compared to determinate which IMF can be used for reconstruction. In another study [41], the authors proposed a noise removal method based on multi-level singular value decomposition and compressed sensing.

- Segmentation and location

Segmentation is one of the faced difficulties in the PCG classification. The key step in PCG signal analysis is to estimate the independent cardiac cycle, which is based on PCG segmentation. The aim of segmentation is to detect the location of the fundamental heart sounds, including the first (S1) and second (S2) heart sounds, which are the important physiological characteristics of heart sounds. Afterwards, the cardiac cycle can be estimated according to the location of S1 and S2. Most of segmentation techniques use Electrocardiogram (ECG) [32]. However, PCG and ECG recordings are not available simultaneously. A complete analysis of the techniques and trends in localization of heart sounds has been provided in [6]. One of the popular PCG segmentation approaches is based on envelope extraction after chosen transform [19, 9, 36]. The envelope function of an oscillating signal is defined as a smooth curve outlining its extremes. In [23], the authors calculated the Shannon energy envelope after applying the S-transform on PCG signal for segmentation, while [35] used the same Shannon envelope but based on the empirical wavelet transform.

- Feature extraction and classification

Feature extraction is a key phase in PCG analysis. Through various techniques, original PCG signals can be transferred to feature vectors with low dimension and meanwhile information remained. The features on PCG signals mainly can be extracted in three ways: time domain [12], frequency domain [29], and time-frequency domain [21, 2]. Although some important information of PCG signals can be found in time domain, there also has discriminative information in frequency domain that can not be ignored. These features are commonly extracted by using transform ways, such as short time Fourier transform [31], S transform [23] and wavelet transform[2, 4]. EMD is another signal decomposition technique and can decompose a PCG signal into several IMFs which contain different frequency components of the original signal, and then one can extract the discriminative features [3, 26]. After extracting features, the next step is to select to classifier to complete the PCG classification. Researchers have used various machine learning as well as non-machine learning algorithms to perform the classification, such as ANN [34], SVM [10, 15, 33] and deep learning [8, 25, 28]. In this paper, we choose SVM as the classifier.

65 This paper is organized as follows: In Section 2, the proposed method is described in detail. It will be seen that a novel envelope extraction model is provided. Based on the proposed envelope model, cardiac cycles of heart sounds can be estimated and then the time-domain features of PCG signals are extracted. The features in frequency domain and wavelet domain are also obtained in this section. Results and performance evaluation are presented in Section 3. Finally, Section 4 concludes this paper.

70 2. Method

In this section, we describe the proposed model for the classification of heart sound signals. Our method can be boiled down to the following steps:

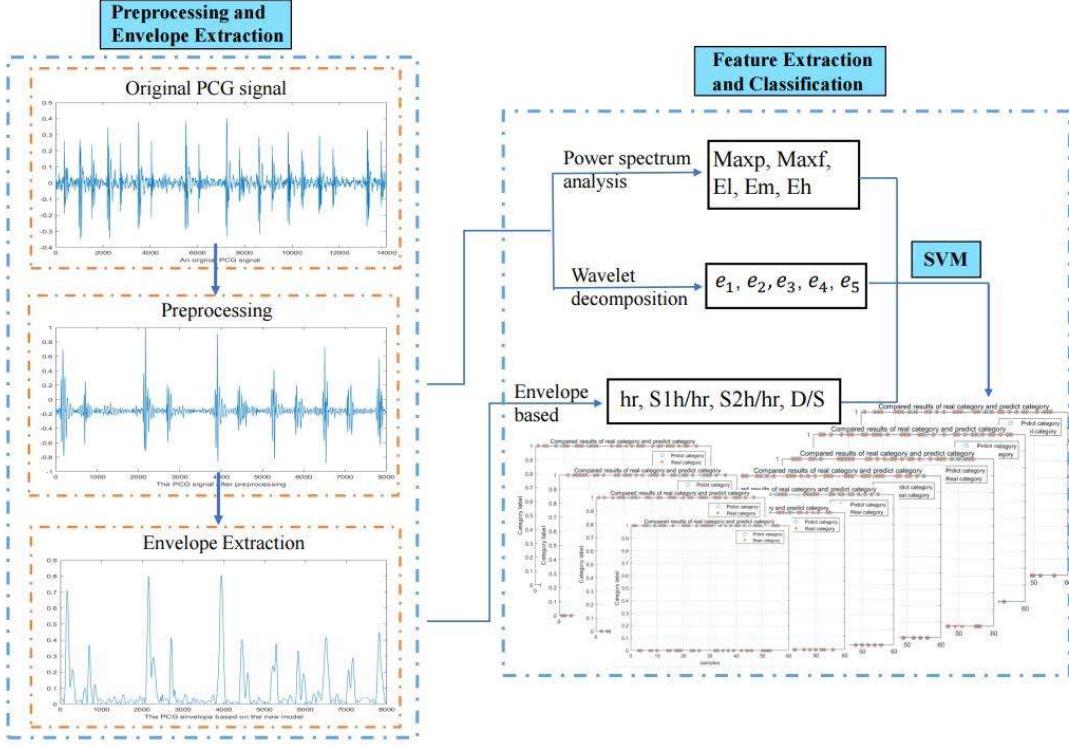


Figure 1: A brief illustration of the proposed model.

Step 1 Preprocessing. The collected heart sound signals are always contaminated with high frequency noise. Most information in heart sounds is contained in the low frequency components. Therefore, the filter process is needed in preprocessing. Moreover, interception and normalization are also necessary in order to get the signals with the same length and intensity.

Step 2 Cardiac cycle estimation based on envelope extraction. Cardiac cycle is important in PCG classification, which is commonly estimated by envelope extraction. A novel envelope model based on constrained optimization will be proposed in this paper. We do not extract the envelope from PCG signal directly, instead we extract the envelope of the chosen IMF of each PCG signal after using EMD method. According to the envelope, S1 and S2, the most important components of heart sound, are identified to divide the heart cycle, which is used for further feature extraction.

Step 3 Feature extraction. In this paper, the features of PCG signals derive from three aspects: time-domain features based on envelope which contain the heart rate, the ratio of S1 duration to heart rate, the ratio of S2 duration to heart rate, the ratio of diastole to systole, wavelet-domain features based on wavelet transform which contain wavelet coefficient energies, frequency-domain features based on Fourier transform which contain the energy proportions of three frequency bands, the maximum power spectral density and the corresponding point.

Step 4 Classification. Feature extraction is followed by the classification of heart sound signals using SVM, which has a linear kernel.

A brief illustration of our method is given in Fig. 1.

2.1 Preprocessing

In the first step, PCG signals are preprocessed for further noise reduction. The main information is concentrated at low frequencies. Therefore, all the PCG signals are resampled to 2000 Hz and Butterworth band pass filter with order 5 and cutoff frequencies from 30 HZ to 200 Hz is applied. Additionally, wavelet threshold denoising method is used for removing the noise which spectrum is overlapped with those of PCG signal. In this

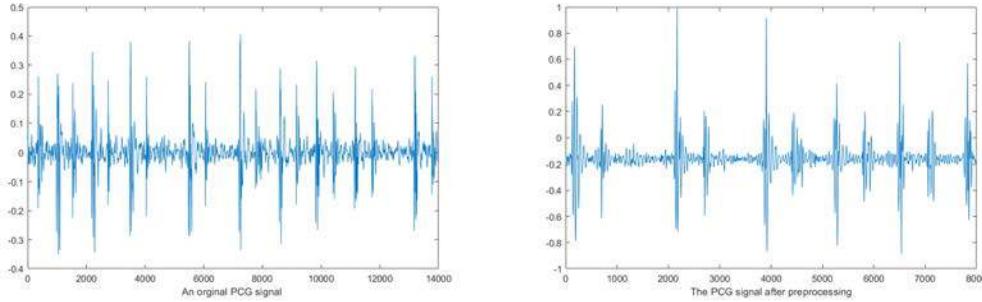


Figure 2: Before and after preprocessing of a PCG signal.

paper, we choose sym8 as the wavelet basis function to perform 3 layer decomposition by using the Rigrsure soft threshold. Then, the PCG signals are normalized to a fixed scale of $[-1, 1]$:

$$y[n] = \frac{x[n]}{\max_n(|x[n]|)} \quad (1)$$

where $x[n]$ is the heart sound signal before normalized and $y[n]$ is the normalized signal.

100 Due to the synchronization with the biological process of repeating the heart cycle, heart sound signal is approximately periodic. Therefore, we can apply the interception method, which can not only reduce the complex of data processing, but also obtain a relatively complete cardiac cycle. The detailed process of interception are as follows:

- 105 • A suitable threshold α is selected, and the normalized signal is binarized according to this threshold. Suppose $z[n]$ is the signal after binarized, then we have:

$$z(n) = \begin{cases} 1, & |y[n]| \geq \alpha \max_n(y[n]) \\ 0, & |y[n]| < \alpha \max_n(y[n]). \end{cases} \quad (2)$$

- 110 • Since the sampling frequency of the signal in this paper is 2000 Hz, find out where the first two thousand points have the largest interval, and use the position corresponding to half of the interval as the starting point for interception. From the starting point, intercept 8000 sample points backwards, and use these 8000 sample points as the last intercepted signal.

115 The heart sound signal before and after interception processing is shown in Fig. 2.

2.2 Heart sound signal envelope extraction

Cardiac cycle is important in heart sound classification, which is commonly estimated by envelope extraction. In this subsection, we propose a novel envelope model based on constrained optimization. We do not extract the envelope from heart sound signal directly, instead we extract the envelope of the chosen IMF after using 115 EMD. The IMFs of each PCG signal are separated by the iterative sifting process of EMD which arranged by frequency from high to low. The frequency of the main components of the heart sound signal is concentrated in the range of 50-200 Hz, which is in the middle of the entire signal frequency range. Therefore, in most cases, the second or third IMF is selected as the best IMF component that represents the corresponding PCG signal. In this paper, the second IMF of each PCG signal is chosen for further analysis. For detailed description of EMD, the interested readers are referred to [18].

120 In our previous paper, we proposed an optimization model of signal envelope [37]. Since the second IMF of each heart sound signal is chosen to estimate the cardiac cycle, the envelope of the IMF should approximately satisfy that the upper and lower envelopes are symmetric. Let $s[n] = |z[n]|$, assuming $u[n]$ is the upper envelope of signal $s[n]$. According to [37], $u[n]$ can be calculated by the following L_1 -minimization problem:

$$\begin{cases} \text{Minimize}_{u \in \mathbb{R}^n} & \|\Phi u\|_1; \\ \text{Subject to} & u \geq s, Au = b. \end{cases} \quad (3)$$

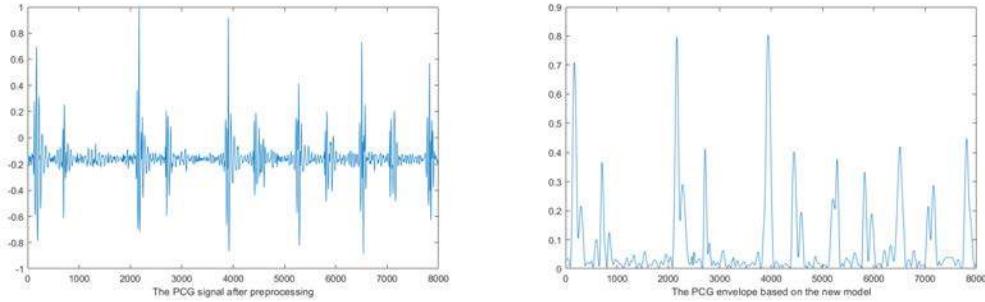


Figure 3: A normal PCG signal and its envelope.

where Φ is 4th-order difference matrix:

$$\Phi = \begin{pmatrix} 1 & -4 & 6 & -4 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & -4 & 6 & -4 & 1 & 0 & \cdots & 0 \\ \cdots & \cdots \\ 0 & \cdots & 0 & 1 & -4 & 6 & -4 & 1 & 0 \\ 0 & \cdots & \cdots & 0 & 1 & -4 & 6 & -4 & 1 \end{pmatrix}.$$

- 125 A is the position matrix of order $M \times n$ and b is the extrema vector of order M , which all are determined by the tangential points. The calculation of A and b can be referred to [37]. Problem (3) can be solved by Algorithm 1 which is based on the split Bregman iteration.

Algorithm 1. Upper Envelope Computation Algorithm

1. initialize: $b^0 = c^0 = d^0 = 0, u^0 = s;$
130 2. while not converge
 - $b^{k+1} = b^k + (b - Au^k), \quad c^{k+1} = c^k + (\Phi u^k - d^k);$
 - $u^{k+1} \leftarrow \min\{\frac{\lambda}{2} \|Au - b^{k+1}\|^2 + \frac{\mu}{2} \|d^k - \Phi u - c^{k+1}\|^2\};$
 - $u^{k+1} \leftarrow \text{Proj}(u^{k+1});$
 - $d^{k+1} \leftarrow \min\{\|d\|_1 + \frac{\mu}{2} \|d - \Phi u^{k+1} - c^{k+1}\|^2\};$
135 3. end while.
-

Fig. 3 illustrates the envelope of one normal PCG signal.

2.3 Segmentation of Heart sound signals

- Once obtaining the envelopes of PCG signals, we can estimate their cardiac cycles and segment the PCG signals consequently. In this subsection, we use the method in [40] for PCG segmentation. This method estimates the cardiac cycle in advance, meanwhile adjusts the errors during the segmentation process, which makes that errors in the segmentation process of one-stop cardiac record will not affect the accuracy of subsequent segmentation. Different from [40], we use the proposed envelope model instead of Shannon energy envelope.

Suppose the length of preprocessed PCG signal $z[n]$ is N , and the envelope of $z[n]$ is $u[n]$, which can be calculated by the following formula

$$C[m] = \sum_{n=1}^{N-n+1} u[n] \frac{u[n+m-1]}{N-n+1}, \quad (4)$$

- 145 where $N - n + 1$ is the equilibrium factor in order to avoid the autocorrelation coefficient $C[m]$ reduces rapidly with the m increasing. Then divide $C[m]$ into several subsequences. Then binary processing is carried out for each subsequence. After that, calculate the mean and variance of the adjacent peaks of $C[m]$. Get rid of the mean which has a big variance, then the average of the remaining mean values can be used as the estimation of cardiac cycle T . With T as a reference, we can segment the PCG signals using the following steps:

- 150 (1) Set the segmentation threshold $Th = \beta * \text{mean}(u)$, where β is a parameter and in this paper $\beta = 0.9$.
Based on the threshold Th , all the rising and falling segmentation points $Tr = [Tr_1, Tr_2, \dots, Tr_m]$ and $Tn = [Tn_1, Tn_2, \dots, Tn_m]$ of the envelope u can be extracted. The area enclosed by the envelope u between each pair of rising and falling segmentation points is calculated. If the area is too small, the corresponding segmentation points are discarded.
- 155 (2) Extract the peak points $P_i, i = 1, 2, \dots, m - 1$ between each pair of Tr_i and Tn_i . Calculate the time intervals $P_{i+2} - P_i, i = 1, 2, \dots, m$ sequentially, if $P_{i+2} - P_i \ll T$, then discard P_i . If $P_{i+2} - P_i \gg T$, then reduce Th , repeat the step (1). After adjusting peak points, we calculate the time difference between the remaining peaks $P_{j+1} - P_j, j = 1, 2, \dots, M - 1$, and recognize the systolic or diastolic duration based on the fact that the systolic duration is smaller than diastolic duration.

160 **2.4 Feature extraction and classification**

After segmentation, time features can be extracted from the PCG signals. In this paper, the time features contain that the heart rate hr , the ratio of S1 duration to heart rate $S1h/hr$, the ratio of S2 duration to heart rate $S2h/h$, the ratio of diastole to systole D/S . Beside these time features, we also extract the features in frequency domain and wavelet domain. The 4th-order Daubechies wavelet has been used in this paper, and the energies of the 4 level detail wavelet coefficients and approximate coefficients are extracted as the wavelet domain features of PCG signals, that is e_1, e_2, e_3, e_4, e_5 . Meanwhile, we can divide PCG signals into three frequency bands: high frequency, medium frequency and low frequency, and calculate the energy proportions between each frequency band and the total energy, which are Eh, Em, El . In addition, calculate the maximum power spectral density Maxf and the corresponding point Maxp . All these features of each PCG signal make up a 14-dimensional feature vector

$$[hr, S1h/hr, S2h/hr, D/s, e_1, e_2, e_3, e_4, e_5, \text{Maxp}, \text{Maxf}, Eh, Em, El].$$

Support vector machine (SVM) has been mostly used in literature for binary classification via non-probabilistic supervised learning model, therefore we use SVM classifier with linear kernel to recognize the normal and abnormal PCG signals in this paper.

3. Performance evaluation

165 In this section, we will test the effectiveness of the proposed model. We first give two artificial signals to test the effectiveness of the proposed envelope model since the true envelopes of the two signals are known. Then this envelope model is used to segment the PCG signals to extract features and the classification results on public datasets are also given. All experiments in this paper are carried out using MATLAB R2016a on a 3.30GHz machine with 4.00GB RAM.

170 **3.1 Effectiveness test of the proposed envelope model**

We give two artificial signals to test our envelope model. The first signal is as follows:

$$x_1(t) = t \sin(t^2), t \in [1, 10]. \quad (5)$$

Obviously, $x_1(t)$ is a mono-component whose envelope is $\rho_1(t) = t, t \in [1, 10]$. The second signal is

$$x_2(t) = (2 + \cos t) \sin(t^2), t \in [5, 15]. \quad (6)$$

$x_2(t)$ is also a mono-component and its envelope is $\rho_2(t) = 2 + \cos t, t \in [5, 15]$. We can use the proposed envelope model to obtain the estimated envelopes of $x_1(t)$ and $x_2(t)$, which are recorded as $u_1(t)$ and $u_2(t)$, respectively.
175 Fig. 4 plots the results of the comparison between the true envelopes and the estimated envelopes. The two subfigures in bottom of Fig. 4 give the error curve $u_i(t) - \rho_i(t), i = 1, 2$. It can be seen that the approximate accuracy reaches more than 10^{-3} which shows the performance of the proposed model. Meanwhile, the following root mean square (RMS) is used to measure quantitatively the error between two signals $u(t)$ and $\rho(t)$:

$$\text{RMS}(u, \rho) = \sqrt{\frac{1}{N} \sum_{i=1}^N |u(t_i) - \rho(t_i)|^2}. \quad (7)$$

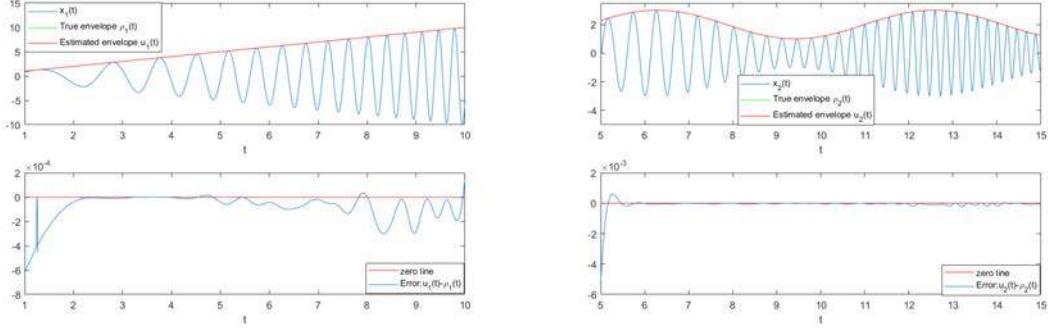


Figure 4: Left: the signal $x_1(t)$, the true envelope $\rho_1(t)$, the estimated envelope $u_1(t)$ and the approximate error $u_1(t) - \rho_1(t)$. Right: the signal $x_2(t)$, the true envelope $\rho_2(t)$, the estimated envelope $u_2(t)$ and the approximate error $u_2(t) - \rho_2(t)$.

It is obvious that the smaller the RMS value, the closer the two signals are. The RMS values between the estimated envelopes and the true envelopes of $x_1(t)$ and $x_2(t)$ are 1.3663×10^{-4} and 3.4764×10^{-4} , respectively, which also shows the performance of the proposed model.

3.2 PCG Classification

3.2.1. Data collection

In this study, two datasets are used to test the effectiveness of the proposed model. Dataset I comes from the database A of the classifying heart sounds PASCAL challenge competition [5], which contains 31 normal and 34 murmurs with a 441000 Hz sampling frequency. These signals all were gathered from the general public via the iStethoscope Pro iPhone app. These signals are of varying lengths, between 1 second and 30 seconds. Dataset II comes from the database A of the competition in 2016 [11], which contains 117 normal and 292 abnormal PCG recordings with 2000 Hz sampling frequency. Similar with Dataset I, the signals in Dataset II is varied from 5 seconds to 120 seconds. As the reference [16], the signals with length less than 7 seconds were excluded. As a result, Dataset I contains 26 normal and 30 murmurs, and Dataset II contains 116 normal and 290 abnormal PCG signals in this study.

3.2.2. Experimental results

In this subsection, classification experiments are performed. For evaluation, we adopt the well-known performance metrics such as sensitivity (S_e), specificity (S_p) and accuracy (AC). Most of the state-of-the-art methods for PCG classification also employ these metrics. The definitions of these metrics are given below:

$$S_e = \frac{TP}{TP + FN}, \quad (8)$$

$$S_p = \frac{TN}{FP + TN}, \quad (9)$$

$$AC = \frac{TP + TN}{TP + FP + FN + TN}, \quad (10)$$

where S_e is the sensitivity, S_p is the specificity, AC is the classification accuracy, TP and TN represent the total number of detected true positives and true negatives, respectively. FP is false positive while FN is false negative. Furthermore, there is usually a trade-off between the sensitivity and specificity. This tradeoff can be represented graphically using a receiver operating characteristic curve (ROC), which is a graphical plot that illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied. AUC is just the area under the curve of ROC and is also suitable to binary classification problem.

Envelope based method is one of important ways used for this task, in which Hilbert envelope and Shannon energy envelope are the most commonly used. Fig. 5 shows the comparison of different envelope based methods for classification. In Fig. 5, we give the ROC curves and AUC values based on different envelope methods. The green line and blue line plot the ROC curves by using Hilbert envelope and Shannon envelope, respectively. They both extract the time features from PCG signals directly without using EMD method first. It can be seen that

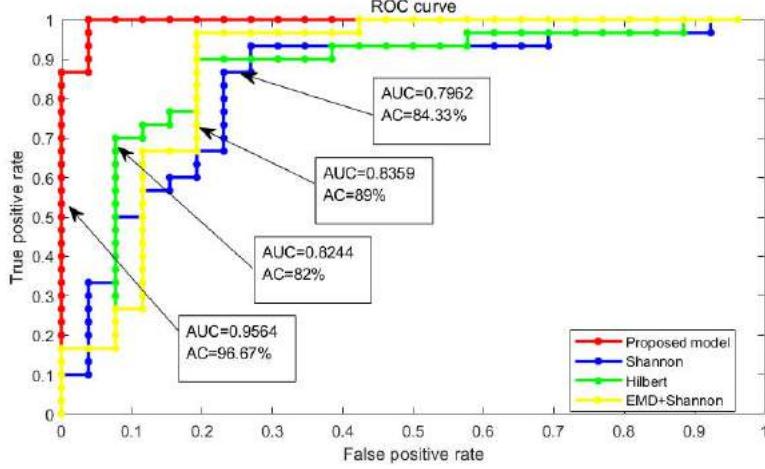


Figure 5: ROC curves comparison of different methods.

Table 1: Results of 10-fold cross validation on two datasets.

fold iteration	Dataset I			Dataset II		
	S_e	S_p	AC	S_e	S_p	AC
1	1	0.6667	0.8333	1	1	1
2	1	1	1	0.9167	1	0.9756
3	1	1	1	1	1	1
4	1	0.6667	0.8333	1	1	1
5	1	1	1	1	1	1
6	1	1	1	1	1	1
7	1	1	1	0.9167	1	0.9756
8	1	1	1	1	1	1
9	1	1	1	1	1	1
10	1	1	1	1	1	1
mean	1	0.9333	0.9667	0.9833	1	0.9951

210 the classification accuracy based on Shannon envelope is higher than that based on Hilbert envelope. Therefore, we choose Shannon envelope plus EMD process to compare with the proposed model. Firstly, we apply EMD process on the each PCG signal and choose the second IMF of it for further analysis. Secondly, we segment these IMFs to estimate the cardiac cycles and then extract the time features. Thirdly, the time features combined with Fourier features and wavelet energy features are fed to SVM for classification. In Fig. 5, the yellow line represents
215 the ROC curve based on Shannon envelope plus EMD process and the red one represents the ROC curve and AUC based on the proposed model. Although compared with using Shannon envelope alone, the accuracy based on Shannon envelope plus EMD process is improved and reaches 89%. It still is lower than the result based on the proposed model which classification accuracy reaches 96.67%.

220 In addition, we give the classification results of 10-fold cross-validation on the two datasets. The details are shown in Table 1. It can be seen that the sensitivities of each fold iteration of Dataset I and the specificities of each fold iteration of Dataset II reach the superior value 1. In both Dataset I and Dataset II, the average accuracies for PCG classification are more than 96%. Specifically, the average accuracy on Dataset I is 96.67% and the average accuracy on Dataset II is 99.51%.

225 We also compare the classification results based on the proposed model with the latest results for PCG classification, which is shown in Table 2. Time-frequency domain features are used in [40]. Meanwhile, its time-domain features contain the electrocardio recordings. The data used in [40] comes from two sources. One is the database A of [5], the other is collected by themselves. They choose the PCG signals which lengths are all more than 10 seconds to the final classification. Finally, they reached 96.08% classification accuracy. The authors in [38] proposed a scaled spectrogram and tensor decomposition based method to extract more discriminative
230 features. They used the data from [11] and SVM classifier. The final accuracy for PCG classification is 90%. [17]

Table 2: Comparison of proposed method with previous methods(%).

	year	Features	Classifier	S_e	S_p	AC
Zhao[40]	2017	Time-frequency domain	SVM	92.31	100	96.08
Zhang[38]	2017	Based on tensor decomposition	SVM	-	-	90
Han[17]	2018	MFCC map	CNN	98.33	84.67	-
Noman[25]	2018	Raw (norm-dur) & MFCC	Ensemble CNN	89.94	86.35	89.22
This paper	2019	Time-frequency domain	SVM	89.83	95.63	96.67
		& wavelet domain		98.33	100	99.51

and [25] both use Mel-frequency cepstrum coefficients (MFCC) for feature extraction and convolutional neural network (CNN) as the classifier. They also used the data from [11]. In [25], the authors combined 1D-CNN and 2D-CNN for classification and reached 89.22% accuracy. However, [25] only used 1D-CNN and reached the sensitivity of 98.33%, the specificity of 84.67%. Based on the proposed model, we achieve the classification accuracy of 96.67% on Dataset I and 99.51% on Dataset II. However, because the experiments in these studies are based on different datasets, it is difficult to directly compare the results in many cases.

4. Conclusion and future work

The main work of this paper is to propose a novel model of envelope extraction. The model is based on the constraint optimization and is effectiveness for mono-components. After using EMD process to each PCG signal, the second IMF is chosen for further analysis as the representation of the corresponding PCG signal. Based on the proposed envelope model and EMD process, the cardiac cycles of PCG signals can be estimated and then the time-domain features can be extracted. Combining with the frequency-domain features and wavelet-domain features, the feature vectors are obtained. Finally, SVM classifier is used to classify the normal and abnormal PCG signals. The proposed framework is evaluated on the public datasets from the PASCAL classifying heart sounds challenge and 2016 PhysioNet challenge. The classification accuracies are both above 96%, which show the effectiveness of the proposed model.

In future work, the fast iterative algorithm will be taken into consideration to improve the speed of the proposed envelope model, which make the big data classification available. Moreover, we will consider the internal and external noise of the PCG signals to improve the robustness of the proposed model. And more efficient and accurate cardiac cycle detection method is required to improve the performance.

Acknowledgements

This research was partially supported by NSFC (No.11701144, 11471101), Higher Education School Young Backbone Teacher Training Program of Henan Province (No.2017GGJS020), Henan Provincial Department of Education Science and Technology Research Key Projects (No.16A120002), NSF of Henan Province (No.162300410061).

References

- [1] M. Ali, E. El-Dahshan, A. Yahia, Denoising of Heart Sound Signals Using Discrete Wavelet Transform, Circuits System Signal Processing 36(2017)4482-4497.
- [2] S. Ari, K. Hembram, G. Saha, Detection of cardiac abnormality from PCG signal using LMS based least square SVM classifier, Expert Systems with Applications, 37(12)(2010)8019-8026.

- [3] S. Ari, G. Saha, Classification of heart sounds using empirical mode decomposition based features, International Journal of Medical Engineering and Informatics, 1(2008)91-108.
- [4] C. Balili, M. Sobrepina, P. Naval, Classification of heart sounds using discrete and continuous wavelet transform and random forests, 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR), DOI: 10.1109/ACPR.2015.7486584.
- [5] P. Bentley, G. Nordehn, M. Coimbra, S. Mannor, The PASCAL Classifying Heart Sounds Challenge 2011 (CHSC2011) Results, <http://www.peterjbentley.com/hearthchallenge/index.html>.
- [6] T. Chakrabarti, S. Saha, S. Roy, I. Chel, Phonocardiogram signal analysis - practices, trends and challenges: a critical review, 2015 International Conference and Workshop on Computing and Communication (IEMCON), Vancouver, BC, Canada, 2015.
- [7] T. Chen, S. Yang, L. Ho, et al., S1 and S2 heart sound recognition using deep neural networks, IEEE Transactions on Biomedical Engineering, 64(2)(2017)372-380.
- [8] W. Chen, R. Gao, L. Liu, M. Zhu, et al., Summary of China cardiovascular disease report 2017, Chinese Circulation Journal, 33(1)(2018)1-8.
- [9] S. Choi, Z. Jiang, Comparison of envelope extraction algorithms for cardiac sound signal segmentation, Expert Systems with Applications 34(2008)1056-1069.
- [10] E. Comak, A. Arslan, İ. Türkoğlu, A decision support system based on support vector machines for diagnosis of the heart valve diseases, Computers in Biology and Medicine, 37(2007)21-27.
- [11] Classification of Normal/Abnormal Heart Sound Recordings, the PhysioNet/Computing in Cardiology Challenge (2016), <https://www.physionet.org/challenge/2016/#challenge-data>.
- [12] S. Deng, J. Han, Towards heart sound classification without segmentation via autocorrelation feature and diffusion maps, Future Generation Computer Systems, 60(2016)13-21.
- [13] N. Dewangan, R. Potda, Noise cancellation using adaptive filter for PCG signal, International Journal of Emerging Trends of Technology in Computer Science, 3(4)(2014)38-43.
- [14] X. Guo, H. Lin, S. Xiao, Medical parameters extraction of heart sounds, Computer Engineering and Applications, 47(3)(2011)214-217.
- [15] G. Güraksin, H. Uğuz, Classification of heart sounds based on the least squares support vector machine, International Journal of Innovative Computing, Information and Control, 7(12)(2011)7131-7144.
- [16] M. Hamidi, H. Ghassemian, M. Imani, Classification of heart sound signal using curve fitting and fractal dimension, Biomedical Signal Processing and Control, 39(2018)351-359.
- [17] W. Han, Z. Yang, J. LU and S. Xie, Supervised threshold-based heart sound classification algorithm, Physiological Measurement, 39(2018)115011(11pp).
- [18] N. E. Huang, Z. Shen, S. R. Long, et al., The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, Proceedings of the Royal Society of London A 454(1971)(1998)903-995.
- [19] Z. Jiang, S. Choi, A cardiac sound characteristic waveform method for in-home heart disorder monitoring with electric stethoscope, Expert Systems with Applications 31(2006)286-298.
- [20] “Cardiovascular diseases (CVDs): Fact Sheets, details” World Health Organization, 17 May 2017. [Online]. Available: [http://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](http://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)).
- [21] I. Maglogiannis, E. Loukis, E. Zafiroopoulos, et al., Support Vectors Machine-based identification of heart valve diseases using heart sounds, Computer Methods and Programs in Biomedicine, 95(1)(2009)47-61.
- [22] S. Messer, J. Agzarian, D. Abbott, Optimal wavelet denoising for phonocardiograms, Microelectronics Journal 32(2001)931-941.

- [23] A. Moukadem et al., A robust heart sounds segmentation module based on S-transform, *Biomedical Signal Processing and Control*, 8(2013)273-281.
- [24] Q. Mubarak, M. Akram, A. Shaukat, F. Hussain, S. Khawaja, W. Butt, Analysisi of PCG signals using quality assessment and homomorphic filters for localization and classificaiton of heart sounds, *Computer Methods and Programs in Biomedicine*, 164(2018)143-157.
- [25] F. Noman, C. Ting, S. Salleh and H. Ombao, Short-segment heart sound classification using an ensemble of deep convolutional neural networks, arXiv:1810.11573v1 [cs.SD] 27 Oct 2018.
- [26] C. Papadaniil, L. Hadjileontiadis, Efficient heart sound segmentation and extraction using ensemble empirical mode decomposition and kurtosis feature, *IEEE Journal of Biomedical and Health Informatics*, 18(4)(2014)1138-1152.
- [27] X. Quan, J. Seok, D. Bae, Detection of S1/S2 components with extraction of murmurs from phonocardiogram, *IEICE Transactions on Information and Systems*, E98-D(3)(2015)745-748.
- [28] J. Rubin, R. Abreu, A. Ganguli, et al., Classifying heart sound recordings using deep convolutional neural networks and mel-frequency cepstral coefficients, *2016 Computing in Cardiology Conference (CinC)*, DOI: 10.23919/CIC.2016.7868867.
- [29] F. Safara, S. Doraisamy, A. Azman, et al., Multi-level basis selection of wavelet packet decomposition tree for heart sound classification, *Computers in Biology and Medicine*, 43(10)(2013)1407-1414.
- [30] A. Salman, N. Ahmadi, R. Mengko, A. Langi and T. Mengko, Empirical Mode Decomposition (EMD) Based Denoising Method for Heart Sound Signal and Its Performance Analysis, *International Journal of Electrical and Computer Engineering*, 6(5)(2016)1-8.
- [31] Y. Soeta, Y. Bito, Detection of features of prosthetic cardiac valve sound by spectrogram analysis, *Applied Acoustics*, 89(2015)28-33.
- [32] D. Springer, L. Tarassenko, G. Clifford, Logistic regression-HSMM-based Heart Sound Segmentation, *IEEE Transactions on Biomedical Engineering*, 63(4)(2016)822-832.
- [33] H. Tang, Z. Dai, Y. Jiang, T. Li, C. Liu, PCG classification using multidomain features and SVM classifier, *BioMed Research International*, 2018, Article ID 4205027, 14 pages.
- [34] I. Turkoglu, A. Arslan, E. Ilkay, An expert system for diagnosis of the heart valve diseases, *Expert Systems with Applications*, 23(2002)229-236.
- [35] W. Varghees, K. Ramachandran, Effective heart sound segmentation and murmur classification using empirical wavelet transform and instantaneous phase of electronic stethoscope, *IEEE Sensors Journal*, 17(12)(2017)3861-3872.
- [36] Z. Yan, Z. Jiang, A. Miyamoto, Y. Wei, The moment segmentation analysis of heart sound pattern, *Computer Methods and Programs in Biomedicine* 98(2010)140-150.
- [37] L. Yang, Z. yang, F. Zhou, L. Yang, A novel envelope model based on convex constrained optimization, *Digital Signal Processing*, 29(2014)138-146.
- [38] W. Zhang, J. Han, S. Deng, Heart sound classification based on scaled spectrogram and tensor decomposition, *Expert Systems With Applications*, 84(2017)220-231.
- [39] W. Zhang, J. Han, S. Deng, Heart sound classification based on scaled spectrogram and partial least squares regression, *Biomedical Signal Processing and Control*, 32(2017)20-28.
- [40] Z. Zhao, X. Zhang, Z. Fang, X. Chen, L. Du, T. Li, Phonocardiogram segmentation and abnormal phonocardiogram screening algorithm based on cardiac cycle estimation, *Journal of Electronics & Information Technology*, 39(2017)2677-2683.
- [41] Y. Zheng, X. Guo, H. Jiang, B. Zhou, An innovative multi-level singular value decomposition and compressed sensing based framework for noise removal from heart sounds, *Biomedical Signal Processing and Control*, 38(2017)34-43.

Few-shot Lifelong Reinforcement Learning with Generalization Guarantees

Anonymous Authors¹

Abstract

In lifelong reinforcement learning (RL), an agent interacts with a sequence of different but related tasks. The goal of lifelong RL is to let the agent quickly adapt to any new task by distilling knowledge from past tasks. Most existing lifelong RL or meta RL algorithms are based on heuristic methods, lacking theoretical guarantees. In this paper, we propose a new lifelong RL approach with theoretical performance guarantee, by extending the PAC-Bayes theory in supervised learning to the regime of RL. We train a distribution of policies, and gradually improve the distribution parameters via optimizing the generalization error bound using trajectories from each task. As the agent sees more tasks, it learns better prior distributions of policies, which results in tighter generalization bounds and further improves future learning. We test our proposed algorithm on various OpenAI's Gym environments. Experimental results show that our proposed algorithms can efficiently adapt to new tasks by continuously distilling knowledge from past tasks and outperforms recent state-of-the-art lifelong RL algorithms.

1. Introduction

Deep reinforcement learning has demonstrated outstanding performance over many challenging tasks, such at playing games, visual navigation, robotic control and manipulation. However, while the current deep RL methods can learn individual policies for specific tasks, it remains very challenging to train a single network that generalizes across different tasks, which is more common in practice. Therefore, we consider the problem of lifelong reinforcement learning, where an agent is facing a series of different but related tasks. We want the policy learnt from the previous tasks can be fast and reliably transferred to new tasks, even in

¹Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

a few-shot learning case where the agent can only interact with each task for a small number of steps.

There are two major challenges for few-shot lifelong learning. (1) It is difficult to confidently determine **what knowledge should be transferred**. For example, if an agent is deployed in a series of mazes, seeking rewards located at random states. The ideal knowledge to be transferred among tasks should be how to walk and how to navigate, instead of the concrete goal locations. Inappropriate knowledge could result in “negative transfer”, which not only does not help, but also harm the learning of future tasks. (2) It is challenging to distill useful knowledge when the samples gained from each task is limited. The agent needs to achieve **fast adaptation** in order to obtain both high rewards and meta knowledge.

Some lifelong learning methods or multi-task learning methods partially address these challenges by providing formal PAC guarantees (Brunskill & Li, 2013; Sun et al., 2020). However, their theory mainly focus on tabular RL, and can not adapt to large-scale environments. For large and complex environments, existing lifelong RL or meta RL works (Finn et al., 2017) use deep neural networks to achieve efficient adaptation. Although good empirical results are reported, these methods are usually based on heuristic, lacking theoretical guarantees.

Therefore, our goal is to design an algorithm that can learn common knowledge from past tasks and quickly adapt to new tasks, as well as provide a reliable theoretical guarantee to avoid negative transfer.

A recent work (Majumdar et al., 2018) propose to optimize a policy distribution over past tasks and directly use it for the following tasks. By applying the well-known PAC-Bayes theory, their algorithm has a nice theoretical guarantee and works for deep neural networks. However, their algorithm can not be used in lifelong learning, where the tasks can be very diverse, making the learnt policy from previous experiences hardly perform well in new tasks. Also, their algorithm and theory are heavily dependent on the selection of a prior policy distribution, and a bad selection of prior may cause significant drop of the performance.

Inspired by Majumdar et al. (2018), we propose an algorithm called PAC-Bayes Lifelong RL (PB-LRL). PB-LRL

055 seeks a *default policy distribution* that generalize well to all
 056 tasks in the underlying tasks distribution, and then for any
 057 new task, we start from a policy sampled from the default
 058 policy distribution. As justified by Finn et al. (2017), a good
 059 policy initialization can significantly improve the learning
 060 performance and achieve fast adaptation. We continuously
 061 update the default policy distribution by repeatedly applying
 062 the PAC-Bayes bound. More importantly, we propose to
 063 “learn” the prior policy distribution required in PAC-Bayes
 064 theory along with the default policy distribution. Thus, we
 065 do not require any expert knowledge and the prior can be
 066 automatically updated to improve the performance.
 067

Our contribution mainly includes the following aspects:

- 070 1. We present a practical algorithms for lifelong reinforcement
 learning with theoretical guarantee, which learns
 a policy distribution that can adapt to new tasks within
 a small number of steps.
- 071 2. Our proposed algorithm can continuously evolve by
 repeatedly utilizing the PAC-Bayes theory to optimize
 the policy distribution and its prior, without any expert
 knowledge of environments.
- 072 3. Experimental results on diverse OpenAI’s Gym environ-
 ments show that our algorithm significantly outper-
 forms state-of-the-art lifelong RL algorithms.

073 2. Related Works

074 **Meta-Learning:** Agents’ lifelong learning requires the generalization
 075 guarantee of policy to new tasks. Meta-learning
 076 has been studied by Bengio et al. (1991), Naik & Mammone
 077 (1992), Thrun & Pratt (1998) and Finn et al. (2017), which
 078 is mostly related to lifelong learning by an efficient meta-
 079 learner. Schmidhuber (1992) and Bengio et al. have trained
 080 a meta-learner that learns how to update the parameters of
 081 the learner’s model. Hochreiter et al. (2001), Andrychowicz
 082 et al., Li & Malik (2017) and Ha et al. (2016) applied this
 083 approach to learn to optimize deep networks for few-shot
 084 learning.

085 Koch, Ravi & Larochelle (2017), Vinyals et al. (2016), Reed
 086 et al. (2018) and Snell et al. (2017) have developed few-shot
 087 methods for supervised-learning. Due to the difference of
 088 problem settings of supervised-learning and reinforcement
 089 learning, which prevents these methods from extension to
 090 reinforcement learning settings.

091 Duan, Wang et al. (2017) has trained memory augmented
 092 recurrent learner to adapt to many tasks, and learned fast
 093 reinforcement learning agents. Saxe et al. (2014), Kirk-
 094 patrick et al. (2017), Krähenbühl et al. (2016), Salimans &
 095 Kingma (2016), Husken & Goericke (2000) and Maclaurin
 096 et al. have considered sensitivity in deep networks, often by

097 good random initialization or data-dependent initialization.
 098 But these methods are often data-inefficient.

099 Finn et al. (2017) directly updated the meta learner weights
 100 using the gradient based on sensitivity on a given task distri-
 101 bution. Their method is agnostic to the form of the model
 102 and to the particular learning task, their methods are comple-
 103 mentary to the approach presented here and could potentially
 104 be used as a baseline for our method. Our experiments show
 105 that our method outperforms their approach.

106 **RL Generalization:** For RL, many works have mentioned
 107 the generalization gap under the policy space for new tasks
 108 and new environments. Early works such Nichol et al.
 109 (2018), Mnih et al. (2013) and Song et al. (2019) have
 110 quantified the generalization of learned policies. An effec-
 111 tive generalization can require an extremely large number
 112 of training environments Cobbe et al. (2019), Cobbe et al.
 113 (2020). And generalization gap always exists in the sim-to-
 114 real problems Tobin et al. (2018), Peng et al. (2018), Tan
 115 et al. (2018).

116 Srivastava et al. (2014), Ioffe & Szegedy (2015), Pacelli
 117 & Majumdar (2020) and Goyal et al. (2019) have applied
 118 regularization approaches to improve the generalization,
 119 but they do not explicitly exploit the structure of the RL,
 120 or limited to a particular task which prevents learner to
 121 exploit causal relationships in the environment. Sinha et al.
 122 (2020) have done adversarial perturbations to the underlying
 123 data distribution to provides robustness guarantees but the
 124 amount of perturbations is not easy to learn.

125 **PAC-Bayes Theory:** In classical supervised learning works
 126 done by Langford & Shawe-Taylor (2002), Seeger (2002),
 127 Germain et al. (2009) and in deep learning works by Dziu-
 128 gaite et al. (2020), Neyshabur et al. (2018) and Neyshabur
 129 et al. (2017), they utilized PAC-Bayes theory McAllester
 130 (1999b) to study the generalization bounds.

131 Neu et al. (2017), Kearns et al., Bagnell & Schneider (2001)
 132 and Bagnell (2004) studied “regularizing” policies to pre-
 133 vent over-fitting and lead to sample efficiency. So Schul-
 134 man, Fard & Pineau (2010) and Fard et al. (2012) have
 135 applied PAC-Bayes theory to learn controlled policies for
 136 Markov Decision Processes with provable sample complex-
 137 ity bounds. However, these approaches required learner
 138 with multiple interactions with a given MDP which is less
 139 efficient.

140 Then Majumdar et al. (2018), Veer & Majumdar (2020)
 141 have made provable generalization guarantees under dis-
 142 tributional shifts and learned policies that could zero-shot
 143 generalized well to novel environments by obtaining upper
 144 bounds on the expected cost of policies on novel environ-
 145 ments . However, they have not shown whether or not the
 146 bound can be workable in multi-task lifelong settings which
 147 in general is more complicated.

In our work, we generated a PAC-Bayes bound as part of the objective of a lifelong learning algorithm to achieve provably data-efficient control on novel tasks. While reusing knowledge from past tasks is a crucial ingredient in making high-capacity scalable models, we focused on the multi-task lifelong RL settings, and we used zero-shot generalization to novel tasks by directly operating on policy space to obtain upper bounds on the expected cost of meta policies, which functioned as a good policy initialization for new tasks.

Like Finn et al. (2017) our method also does not induce additional parameters and uses the same gradient descent update for both learner and meta learner to provide a good initialization to the learner and achieve extremely efficient adaptation to new tasks in only a few gradient steps.

3. Preliminaries

In this section, we discuss the primary technical background that we leverage in this paper.

3.1. Reinforcement Learning

In RL, an agent interacts with the environment by taking actions, observing states and receiving rewards. The environment is modeled by a Markov Decision Process (MDP), which is denoted by a tuple $\mathcal{M} = \langle \mathcal{S}, \mathcal{A}, P, R, \gamma, \mu \rangle$, where \mathcal{S} is the state space, \mathcal{A} is the action space, P is the transition kernel, R is the reward function, $\gamma \in (0, 1)$ is the discount factor, and μ is the initial state distribution.

A trajectory $\tau \sim \pi$ generated by policy π is a sequence $s_1, a_1, r_1, s_2, a_2, \dots$, where $s_1 \sim \mu$, $a_t \sim \pi(a|s_t)$, $s_{t+1} \sim P(s|s_t, a_t)$ and $r_t = R(s_t, a_t)$. The goal of an RL agent is to find an optimal policy π^* that maximizes the *expected total rewards* η , which is defined as $\eta(\pi) = \mathbb{E}_{\tau \sim \pi}[r(\tau)] = \mathbb{E}_{s_1, a_1, \dots, \mu, \pi, P, R}[\sum_{t=1}^{\infty} \gamma^{t-1} r_t]$.

3.2. Meta Learning and Lifelong Learning

In lifelong RL, the agent interacts with a (probably infinite) sequence of tasks, which are i.i.d coming from an underlying distribution of tasks (MDPs), denoted as \mathcal{D} . Suppose that these tasks share the same $\mathcal{S}, \mathcal{A}, \gamma, \mu$, but may have different transition probabilities P and rewards R . The learning process is:

1. initialize a policy π_0 ;
2. sample a task (MDP) $\mathcal{M}_i \sim \mathcal{D}$;
3. starting from π_0 , learn a policy π_i for task M_i to maximize rewards.

3.3. PAC-Bayes Theory in Supervised Learning

Suppose \mathcal{Z} is an input space and \mathcal{Z}' is a set of labels. Let \mathcal{D} be the (unknown) true distribution on \mathcal{Z} . Let \mathcal{H} be a hypothesis class consisting of functions $h_w : \mathcal{Z} \rightarrow \mathcal{Z}'$

parameterized by $w \in \mathbb{R}^d$ (e.g., neural networks parameterized by weights w). Let $l : \mathcal{H} \times \mathcal{Z} \rightarrow \mathbb{R}$ be a loss function. We will denote by \mathcal{P} the space of probability distributions on the parameter space \mathbb{R}^d . Informally, we will refer to distributions on \mathcal{H} when we mean distributions over the underlying parameter space. PAC-Bayes analysis then applies to learning algorithms that output a distribution over hypotheses. Generally, such algorithms will be given a "prior" distribution $P_0 \in \mathcal{P}$ in the beginning and learns a posterior distribution $P \in \mathcal{P}$ after observing training data samples.

Let us denote the training loss associated with the posterior distribution P as:

$$l_S(P) := \frac{1}{N} \sum_{z \sim S} \mathbb{E}_{w \sim P} [l(h_w; z)] \quad (1)$$

and the true expected loss as:

$$l_D(P) := \mathbb{E}_{z \sim \mathcal{D}} \mathbb{E}_{w \sim P} [l(h_w; z)] \quad (2)$$

The following theorem provides a useful upper bound for the true expected loss.

Theorem 1 (PAC-Bayes Upper Bound for Supervised Learning) (Maurer (2004); McAllester (1999a)) For any $\delta \in (0, 1)$, with probability at least $1 - \delta$ over samples $S \sim \mathcal{D}^N$, the following inequality holds:

$$\underbrace{l_D(P)}_{\text{True expected loss}} \leq \underbrace{l_S(P)}_{\text{Training loss}} + \sqrt{\underbrace{\mathbb{D}_{KL}(P||P_0) + \log\left(\frac{2\sqrt{N}}{\delta}\right)}_{\text{"Regularizer}}} / 2N \quad (3)$$

4. Our Methods

In this paper, we propose a guaranteed lifelong RL algorithm that incrementally distills knowledge from previous tasks and quickly adapt to a new task.

4.1. PAC-Bayes Bound for Lifelong RL

Our theoretical learning guarantee is mainly based on the PAC-Bayes theory, which was first applied to RL by Majumdar et al. (2018). We carefully extend the work of Majumdar et al. (2018) to the regime of lifelong RL.

Define Π as the whole policy space parameterized by $\theta \in \mathbb{R}^d$ (θ could be a neural network). Let P be any policy distribution over the policy space Π . Then the *cost* of P evaluated on task \mathcal{M} is defined as

$$C(P; \mathcal{M}) = \mathbb{E}_{\pi \sim P} [C(\pi; \mathcal{M})] = \mathbb{E}_{\pi \sim P} [-\eta_{\mathcal{M}}(\pi)], \quad (4)$$

where $\eta_{\mathcal{M}}(\pi)$ is the total expected reward of policy π in MDP \mathcal{M} .

Then, the expected cost of P over the task distribution \mathcal{D} can be bounded by the following theorem.

Theorem 2 (PAC-Bayes Bound for Lifelong RL) *For any $\delta \in (0, 1)$, over the lifelong task distribution \mathcal{D} and N i.i.d. sampled tasks $\{\mathcal{M}_i\}_{i=1}^N$, for any distribution over policy space P and any prior distribution P_0 , the following inequality holds*

$$\underbrace{\mathbb{E}_{\mathcal{M} \sim \mathcal{D}} C(P; \mathcal{M})}_{\text{True cost}} \leq \underbrace{\frac{1}{N} \sum_{i=1}^N C(P; \mathcal{M}_i)}_{\text{Training cost}} + \sqrt{\underbrace{\mathbb{D}_{KL}(P \| P_0) + \log\left(\frac{2\sqrt{N}}{\delta}\right)}_{2N} \text{Regularizer}} \quad (5)$$

with probability at least $1 - \delta$.

Remarks. (1) Theorem 2 guarantees that the expected loss/cost of the policy distribution over the underlying task distribution is upper bounded by the loss/cost computed on the past tasks and a regularization term.
 (2) Theorem 2 holds for any policy distribution P and any prior distribution P_0 in the policy space Π .
 (3) The regularizer involves the number of training tasks N , and the KL-divergence between P and P_0 . Therefore, the tightness of the bound depends on the selection of prior P_0 , and the number of training tasks.

Majumdar et al. (2018) use the PAC-Bayes bound to learn a policy distribution P over a set of similar but noisy tasks so that the expected cost of the policy on a new task is bounded. However, their method cannot be directly applied to our lifelong setting due to the following two reasons. **(1) Task divergence.** the tasks in lifelong learning can be divergent. For example, task A is to find an apple in the upper-right corner in a maze, and task B is to find an orange in the lower-left in the same maze; although the two tasks share the same transition dynamics, they require totally different policies, so that learning one single policy distribution for all tasks will not work well. This is similar to the first challenging we stated in the introduction about transferring proper knowledge. **(2) Choosing prior.** As pointed out by the remarks above, the tightness of the bound depend on the prior policy distribution P_0 . However, in practice, it is hard to choose a good prior before learning starts. Majumdar et al. (2018) select the prior randomly, which is not optimal, or by expert knowledge, which is usually unrealistic.

Therefore, in the next section, we propose a new algorithm for lifelong RL that overcomes the aforementioned two problems.

4.2. Improve Lifelong Learning with PAC-Bayes Bound

Based on Theorem 2, we introduce our algorithm called PAC-Bayes Lifelong RL (PB-LRL), which achieves efficient knowledge transfer and fast adaptation to novel tasks. Note that PB-LRL is a meta-learning algorithm that be combined with any single-task RL algorithm.

Algorithm 1 PAC-Bayes Lifelong RL (PB-LRL)

```

1: Input: policy dimension  $d$ ; learning rates  $\alpha, \beta$ ; update
   frequency  $N$ ; failure probability  $\delta$ ; the number of steps
   allowed in each task  $T$ ; prior evolving speed  $\lambda$ 
2: Initialize default policy mean and derivation  $\mu, \sigma \in \mathbb{R}^d$ ;
3: Initialize prior policy mean and derivation  $\mu_0, \sigma_0 \in \mathbb{R}^d$ ;
4: Initialize cost gradients  $g_\mu[i] = 0, \forall i = 1, 2, \dots, N$ ,
    $g_\sigma[i] = 0, \forall i = 1, 2, \dots, N$ 
5: for  $i = 1, 2, 3, \dots$  do
6:   Receive a new task  $\mathcal{M}_i \sim \mathcal{D}$ 
7:   Sample  $\epsilon \sim \mathcal{N}(0, I_d)$ 
8:   Set initial policy  $\theta \leftarrow \mu + \sqrt{\sigma} \odot \epsilon$ 
9:   Use  $\theta$  to rollout trajectories  $\mathcal{T}$ 
10:  Compute cost  $c = C(\mathcal{T})$ 
11:  Compute gradients  $g_\mu[i] = \nabla_\mu c, g_\sigma[i] = \nabla_\sigma c$ 
12:  Make a copy of policy  $\theta_i = \theta$ 
13:  for  $t = 1, 2, \dots, T$  do
14:    Optimize  $\theta_t$  for  $\mathcal{M}_i$  with any single-task method
15:  end for
16:  if  $i \bmod N = 0$  then
17:    {Meta Update}
18:     $\mu \leftarrow \mu - \beta(\sum_{i=1}^N g_\mu[i] + \nabla_\mu L_r(\mu, \sigma, \mu_0, \sigma_0, N))$ 
19:     $\sigma \leftarrow \sigma - \beta(\sum_{i=1}^N g_\sigma[i] + \nabla_\sigma L_r(\mu, \sigma, \mu_0, \sigma_0, N))$ 
20:     $\mu_0 \leftarrow (1 - \lambda)\mu_0 + \lambda\mu$ 
21:     $\sigma_0 \leftarrow (1 - \lambda)\sigma_0 + \lambda\sigma$ 
22:  end if
23: end for

```

The key to PB-LRL is to learn a policy distribution P as a policy initializer, called *default policy*. More specifically, we optimize the default policy P with Theorem 2, and whenever there is a new task, we sample a behavior policy from P and use any single-task learning method to fine-tune the policy. Hence, every task gets a “customized” policy, while the common knowledge is shared by the default policy, which solves the task divergence problem. Since in lifelong setting, the tasks are streaming in, we update the default policy every N tasks and estimate the training cost of P by the most recent N tasks.

To remedy the challenge of choosing prior, PB-LRL uses an “evolving” prior rather than a fixed prior. After every N tasks, we update P by minimizing the PAC-Bayes bound evaluated at the current prior P_0 . Then, we update the prior policy distribution by making it closer to the new default policy. Because the default policy is updated to minimize

the expected cost, a fixed prior may result in larger and larger $\mathbb{D}_{KL}(P||P_0)$, then looser and looser PAC-Bayes bound. In contrast, we slowly move the prior towards the default policy by $P_0 = (1 - \lambda)P_0 + \lambda P$, where $\lambda \in [0, 1]$ is a hyper-parameter controlling the moving speed. In this way, even if the prior is initialized with a bad choice, we are still able to find a good prior during learning, and use it to improve the default policy.

The proposed algorithm PB-LRL is illustrated in Algorithm 1, where we use d -dimensional Gaussian distributions as the default and prior policy distribution, i.e., $\mathcal{N}(\mu, \sigma)$ and $\mathcal{N}(\mu_0, \sigma_0)$. Before learning starts, the agent initializes a default policy distribution P and a prior policy distribution P_0 randomly or by domain knowledge (Line 2-3). For every task, the agent first samples a policy $\theta \sim P$ (Line 6-8), estimate the cost of θ (Line 9-11), and then fine-tune the policy with any single-task learning algorithm (Line 13-15). After every N tasks, the agent summarizes its recent experience, and updates default policy P by seeking to minimize the generalization error bound (5) evaluated at the current prior P_0 (Line 18-19). Then, the agent also updates the policy prior by $P_0 = (1 - \lambda)P_0 + \lambda P$ (Line 20-21). Note that the prior P_0 is only used to construct the PAC-Bayes bound and optimize the default policy P .

The cost function $C(\pi, \mathcal{M})$ is approximated by $C(\mathcal{T}) = \frac{1}{|\mathcal{T}|} \sum_{\tau \in \mathcal{T}} r(\tau)$, where \mathcal{T} is the trajectories generated by the current policy. The regularization function L in line 18-19 is defined as

$$L(\mu, \sigma, \mu_0, \sigma_0, N) = \sqrt{\frac{\mathbb{D}_{KL}[\mathcal{N}(\mu, \sigma)||\mathcal{N}(\mu_0, \sigma_0)] + \log \frac{2\sqrt{N}}{\delta}}{2N}}$$

Importance of the update frequency N . The value of hyper-parameter N controls how quickly the meta learner – default policy and prior policy – are updated. The larger N is, the slower the meta learning learns. However, a large N also implies the PAC-Bayes upper bound is tighter and thus the update is more confident. Hence, there is an interesting trade-off between efficiency and accuracy. We empirically evaluate the influence of N in experiments in Section 5.4.

5. Experiments

In this section, we show the performance of our proposed algorithm on various environments for the lifelong learning tasks and compare it with the baseline algorithms.

Environments. We evaluate our proposed PB-LRL in multiple OpenAI’s Gym and Mujoco environments, including discrete control (CartPole, LunarLander) and continuous control (Swimmer). For each environment, we extend the original implementation to the lifelong setting, where the tasks have varying goals or dynamics. The agent needs to

sequentially learn multiple tasks, each for a small number of steps. We introduce the detailed settings and results of the three environments respectively in Section 5.1, 5.2 and 5.3.

Baselines. Our PB-LRL is a meta RL method and can use any single-task RL algorithm as the base learner. For simplicity of comparison, we use Vanilla Policy Gradient (VPG) (Sutton et al., 1998) as the base learner. And we compare our PB-LRL with the following two baselines, both of which use VPG as their base learning algorithm.

- Single-task learning. The Single-task learning baseline is learning each task separately and updating policy based on each tasks separately.
- MAML (Finn et al., 2017). MAML is a state-of-the-art meta-learning method, which uses a heuristic method to optimize the meta policy. By simply aggregating the losses of a batch of tasks after one-step learning and taking gradient w.r.t. the meta policy, MAML aims to find model parameters that are valuable for all tasks.

Note that in both baselines, they treat the policy as a single point in the parameter space rather than a distribution as in our algorithm.

Hyper-parameters. In all experiments, we use a 2-layer MLP with 32 nodes per layer to parametrize the policy. The single-task learning rate α and meta learning rate β are both set to be 10^{-4} . The prior-update-speed parameter λ is set as 0.5. The default policy and the prior policy are both initialized as standard normal distributions. In each lifelong environment, we run the agent on 2000 tasks (or 1000 tasks for Swimmer). In every task, we only run 10 episodes with at most 300 steps per episodes. Note that the number of per-task steps is much less than usually required in traditional learning, so that the agent has to achieve fast adaptation to gain high rewards.

The selection of update frequency N for our algorithm is specified in the captions of figures. Similarly, MAML also updates its meta policy every a batch of tasks; for fair comparison, we set MAML’s update frequency to be the same with N in all experiments.

All experimental results are averaged over 10 random seeds to reduce noise, and we plot the confidence interval by 10% standard deviation. And to increase readability, we smooth all the curves in plots using the exponential moving average algorithm with smoothing parameter 0.99.

5.1. In the CartPole Environment

In the classic CartPole environment depicted in Figure 1(a), a pole is attached by an un-actuated joint to a cart, which moves along a frictionless track. The system is controlled by applying a force of +1 or -1 to the cart. A reward of +1

is provided for every timestep that the pole remains upright. The episode ends when the pole falls or the cart moves far away from the center. We extend the original setting into a lifelong setting by defining two variants of CartPole, as respectively shown in Section 5.1.1 and Section 5.1.2.

5.1.1. TASKS WITH VARYING GOALS.

We implement a variant of CartPole, called CartPole-Goal, where the agent is supposed to reach a “goal position” on the track. The reward is defined by $r(x) = \exp |x - x_{goal}|$, where x is the agent’s current position, and x_{goal} is the goal’s position. In a lifelong setting, the agent interacts with a series of tasks with different goal positions. For every task, the goal position is sampled from a Gaussian distribution centered at 0. The variance/standard deviation controls how divergent those random tasks are. We separately test two different standard deviations: 0.1 and 0.5. Goals with standard deviation 0.1 are more similar with each other, but the goals with standard deviation 0.5 are more divergent and harder to transfer knowledge. (Note that the active position range along the track is $[-2.4, 2.4]$, thus the goals with standard deviation 0.5 are sufficiently spread out over the track.) An ideal agent should distill the knowledge of “how to keep balance” from previous tasks, and quickly adapt to new tasks with different goals.

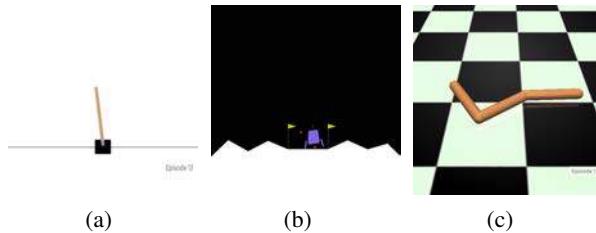


Figure 1. The illustration of three environments. (a) CartPole, (b) LunarLander, (c) Swimmer

A comparison of the performance of our PB-LRL and baselines is shown in Figure 6, on CartPole-Goal environments with standard deviations 0.1 and 0.5. We can see that PB-LRL significantly outperforms MAML and the single-task learner in both case. As it sees more and more tasks, PB-LRL gradually and constantly improves its performance on each of the single tasks. In Figure 6, because of the high divergence of goals, we observe some fluctuation of PB-LRL, but over 2000 tasks, PB-LRL successfully overcomes the divergence of tasks, and achieves fast adaptation to the later tasks. In contrast, the performance of MAML increases much slower than PB-LRL. In Figure 2(b), MAML achieves small progress compared with the naive single-task due to the high divergence of tasks.

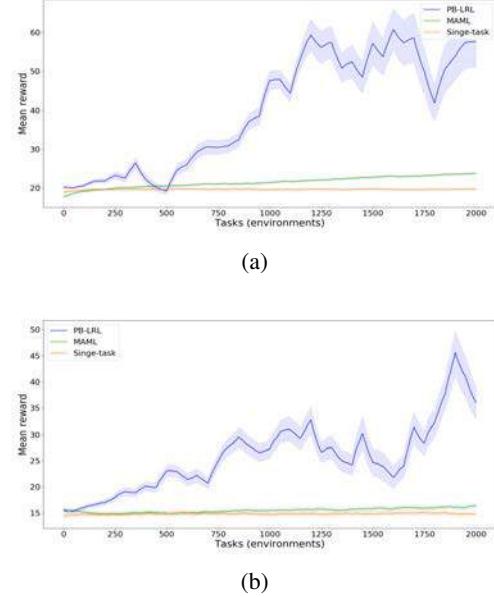


Figure 2. Comparison between our algorithm PB-LRL and baselines on CartPole-Goal. (a) $x_{goal} \sim \mathcal{N}(0, 0.1)$, $N = 25$; (b) $x_{goal} \sim \mathcal{N}(0, 0.5)$, $N = 50$;

5.1.2. TASKS WITH VARYING DYNAMICS.

In this section, we provide experimental results to show our algorithm outperforms baselines when the underlying tasks have different dynamics.

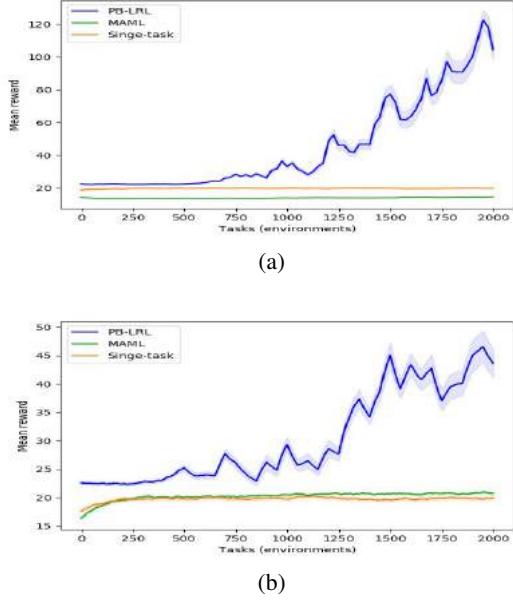
The intrinsic dynamics would affect the lifelong learning performance. For example, in the Cart-Pole environment (Figure 1(a)), different mass of the cart could generate different friction force between the cart and the track, which later affects the acceleration of the cart (the acceleration of the cart is less with a heavier cart) and the rotation of the pole, then it will result in the MDP model alternated so the knowledge transfer between meta learner and learner changed.

To help show that, we compared our algorithm with baselines by changing the single task mass, called CartPole-Mass (m_c). With other conditions of the environment fixed (friction force, controller force, etc.). Specifically, we set $m_c \sim \mathcal{N}(\mu_c, \sigma_c)$, and for every task, we sampled $\{m_{ci}\}_1^{2000}$ from the above distribution.

Figure 3 shows performance of our method with baselines under different single task dynamics when we set $\mu_c = 0.5$ or $\mu_c = 1$. Each single task has different m_i which results in 2000 different small dynamics.

Figure 3(a) and 3(b) show that our method is able to learn the policy that leads to the highest reward compared to baselines. We also see that MAML is more sensitive to single task

330 dynamics. The performance of our method is invariant to
 331 which single task dynamics it encountered, indicating its
 332 generalization ability to divergent dynamics for lifelong
 333 learning.
 334



356 Figure 3. Comparison between our algorithm PB-LRL and base-
 357 lines on CartPole-Mass. The meta learning update frequency $N =$
 358 50 , Different Cart Mass: (a) $\mu_c = 0.5$; (b) $\mu_c = 1.0$;
 359
 360

5.2. In the LunarLander Environment:

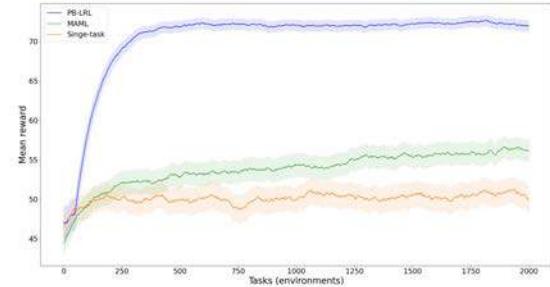
362 The general task in LunarLander environment (as shown in
 363 Figure 1(b)) is to let lunar module land at a pre-defined goal
 364 location.

365 For the LunarLander environment, the agent (lunar module)
 366 needs to learn to land at different goal positions as its tasks.
 367 The goal position in a task is defined as the expected landing
 368 location x_{goal} of the lunar module. We sample uniformly
 369 to get different goals/landing locations. The reward is given
 370 by the distance between the landing position x and goal
 371 position as $r(x) = \exp |x - x_{goal}|$.
 372

373 The Learning progress of PB-LRL is shown as in Figure
 374 4. During the training, the agent is learning from different
 375 goals and samples, our algorithm can still learn to improve
 376 its performance throughout the time. This shows the ability
 377 of our algorithm can learn to quickly adapt to new tasks by
 378 distilling knowledge from previously seen tasks.

379 We also conduct the comparison between our algorithm and
 380 some baselines. As show in Figure 4, we can see that the
 381 MAML performs better than Single-task learning, this is
 382 not surprising since during training MAML takes additional
 383 consideration in performing well on the overall task distri-
 384

bution rather than only on a single task. Our algorithm is
 more efficiently in adapting to new tasks by learning from
 previously seen tasks comparing to the two baselines. The
 possible reason that our algorithm can outperform MAML
 is that we use PAC-Bayes theory in updating our policy.



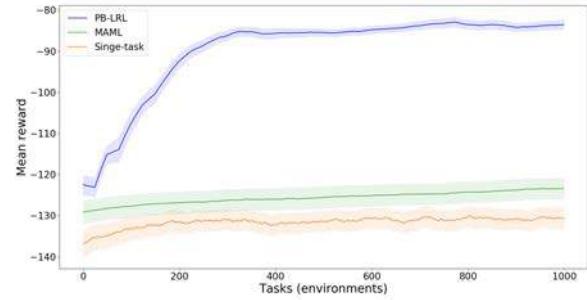
394 Figure 4. Comparison between our algorithm PB-LRL and base-
 395 lines on LunarLander (with 0.1 standard deviation interval). The
 396 meta learning update frequence is $N = 50$.

5.3. In the Swimmer Environment

443 The swimmer is a planar robot with 3 links and 2 actuated
 444 joints. Fluid is simulated through viscosity forces, which
 445 apply drag on each link, allowing the swimmer to move
 446 forward. The goal of the original task is to make the robot
 447 swim forward as fast as possible by actuating the two joints.

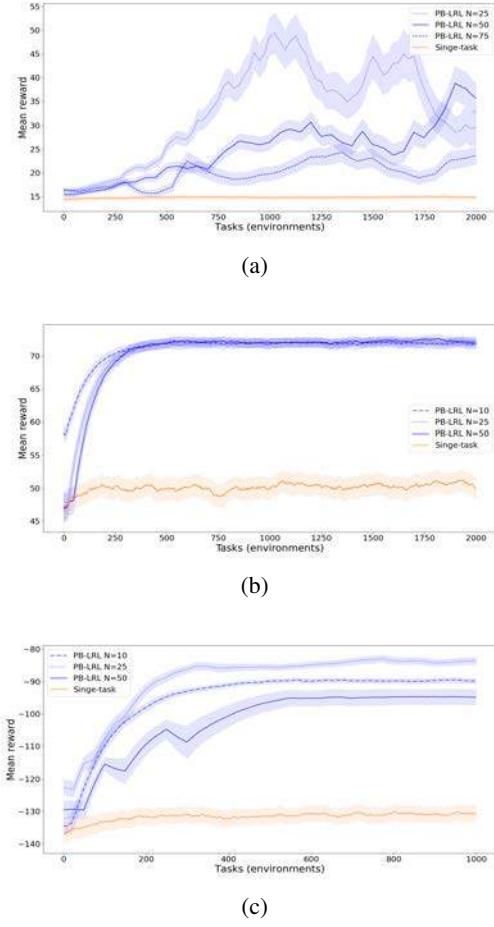
466 In our experiment, we simulate the environments by gener-
 467 ating random goal velocity for the robot, which is uni-
 468 formly chosen from $[0.1, 0.2]$. The reward is defined as
 469 $r(s, a) = -1.5|v_x - v_{goal}| - 1e^{-4}\|a\|_2^2$, where $|v_x - v_{goal}|$
 470 is the difference between the current velocity of the agent
 471 and the goal velocity.

499 We present the comparison results between our algorithm
 500 and the baselines in Figure 5. Our PB-LRL algorithm can
 501 efficiently adapt to new tasks and steadily improve the per-
 502 formance, and constantly outperform MAML and the single-
 503 task baseline.



522 Figure 5. Comparison between our algorithm and baselines on
 523 Swimmer. The meta learning update frequence is $N = 25$.

385 **5.4. Investigating the Effects of N**
 386
 387
 388
 389
 390
 391
 392
 393
 394



417 *Figure 6. Comparison of different update frequency N on (a)*
 418 *CartPole-Goal; (b) LunarLander; (c) LunarLander.*

420 As we discussed in Section 4.2, the hyper-parameter, update
 421 frequency N , balances the trade-off between efficiency and
 422 accuracy of updating policies. In theory, the smaller N leads
 423 to the faster update but will make the learning less stable.
 424 The effect of larger N is a slower but more stable learning.
 425 Our experiment results have verified this theoretical interpretation.
 426 However, from our experiment results, we can also see that the practical effect of N on the performance of
 427 learning also depends on the environments.
 428

429 In the CartPole environment as shown in Figure 6(a), $N =$
 430 25 exhibits the fast and unstable learning behavior as in
 431 theory and $N = 50$ performs the best in the learning w.r.t
 432 final reward. In contrast, in LunarLander (as in Figure 6(b))
 433 and Swimmer (as in Figure 6(c)) environments, for all N
 434 cases, learning curves are nearly stable. In LunarLander all
 435 three N cases reach to similar final rewards, in Swimmer
 436 $N = 25$ reaches to the highest reward. One common thing
 437 is in all three environments, the largest N leads to the slowest
 438 learning, which is also corresponding to the theory.
 439

The interpretation of the dependency on environments for learning behaviors, as shown in our results, can be intuitive. If tasks in a specific environment all share high portion of general features, then the effect of N on stability will be negligible. In addition, in practice when we test the outcome of learning, we only learn from a finite samples of tasks, this actually imposes an implicit requirement on both efficiency and accuracy for learning, which leads to that the moderate N performs the best as in our results.

6. Discussion and Future Works

In this paper, we propose a novel algorithm named PB-LRL for lifelong RL that achieves fast adaptation with theoretical guarantees. We show via systematically designed experiments that PB-LRL significantly outperforms the naive single-task learning algorithm and a state-of-the-art meta-learning algorithm MAML.

We would also like to point out that PB-LRL can be further extended and improved in a number of possible directions, which are exciting to explore in the future.

Automatic and adaptive selection of the meta update frequency N . As we discussed in Section 4.2 and 5.4, the selection of hyper-parameter N is essential for both the theory and the experiment. Moreover, the best selection of N usually depends on the underlying task distributions. In practice, one may need to try different N 's manually to determine which N works better. However, we can get rid of the manual selection by designing an automatic and adaptive algorithm. Our algorithm 1 uses a fixed N , but it also works if N changes after each meta update. More specifically, one can increase N if the learning performance is unstable, and decrease N if the performance is stable but raises slowly. Based on our observation in experiments, it is better to start from a small N and increase it when the performance fluctuates. In this way, we can hopefully converge to a good N and maximize the average rewards.

More gradient descent steps for optimizing default policy. The current PB-LRL uses one-step gradient descent to update the default policy in the ‘meta update’ (Algorithm 1 line 16-22). The gradient step reduces the PAC-Bayes upper bound, but does not necessarily achieve the minimum. If one wants to really minimize the upper bound, more gradient steps are needed. The difficulty of doing so is that we already lost access to the past tasks, thus we cannot directly evaluate the cost of a new default policy on an old task. There are two approximate solutions: (1) using importance sampling to evaluate the new policy with the old trajectories, and (2) building a deep prediction model for each task \hat{M}_i during single-task learning and evaluate new policies on \hat{M}_i , in the scenario where the interactions with each task are sufficient to build a good approximate model.

References

- ## References

Andrychowicz, M., Denil, M., Gómez, S., Hoffman, M. W., Pfau, D., and Schaul, T. Learning to learn by gradient descent by gradient descent. pp. 9.

Bagnell, J. A. *Learning Decisions: Robustness, Uncertainty, and Approximation*. 2004.

Bagnell, J. A. and Schneider, J. G. Autonomous helicopter control using reinforcement learning policy search methods. In *Proceedings 2001 ICRA. IEEE International Conference on Robotics and Automation (Cat. No.01CH37164)*, volume 2, pp. 1615–1620 vol.2, May 2001. doi: 10.1109/ROBOT.2001.932842.

Bengio, S., Bengio, Y., Cloutier, J., and Gecsei, J. On the Optimization of a Synaptic Learning Rule. pp. 29.

Bengio, Y., Bengio, S., and Cloutier, J. Learning a Synaptic Learning Rule. In *In Conference on Optimality in Biological and Artificial Networks*, 1991.

Brunskill, E. and Li, L. Sample complexity of multi-task reinforcement learning. In *Proceedings of the Twenty-Ninth Conference on Uncertainty in Artificial Intelligence, UAI'13*, pp. 122–131, Arlington, Virginia, USA, 2013. AUAI Press.

Cobbe, K., Klimov, O., Hesse, C., Kim, T., and Schulman, J. Quantifying Generalization in Reinforcement Learning. In *International Conference on Machine Learning*, pp. 1282–1289. PMLR, May 2019.

Cobbe, K., Hesse, C., Hilton, J., and Schulman, J. Leveraging Procedural Generation to Benchmark Reinforcement Learning. *arXiv:1912.01588 [cs, stat]*, July 2020.

Duan, Y. Meta Learning for Control. pp. 125.

Dziugaite, G. K., Hsu, K., Gharbieh, W., Arpino, G., and Roy, D. M. On the role of data in PAC-Bayes bounds. *arXiv:2006.10929 [cs, stat]*, October 2020.

Fard, M. and Pineau, J. PAC-Bayesian Model Selection for Reinforcement Learning. *Advances in Neural Information Processing Systems*, 23:1624–1632, 2010.

Fard, M. M., Pineau, J., and Szepesvari, C. PAC-Bayesian Policy Evaluation for Reinforcement Learning. *arXiv:1202.3717 [cs, stat]*, February 2012.

Finn, C., Abbeel, P., and Levine, S. Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks. March 2017.

Germain, P., Lacasse, A., Laviolette, F., and Marchand, M. PAC-Bayesian learning of linear classifiers. In *Proceedings of the 26th Annual International Conference on Machine Learning, ICML '09*, pp. 353–360, New York, NY, USA, June 2009. Association for Computing Machinery. ISBN 978-1-60558-516-1. doi: 10.1145/1553374.1553419.

Goyal, A., Islam, R., Strouse, D., Ahmed, Z., Botvinick, M., Larochelle, H., Bengio, Y., and Levine, S. InfoBot: Transfer and Exploration via the Information Bottleneck. *arXiv:1901.10902 [cs, stat]*, April 2019.

Ha, D., Dai, A., and Le, Q. V. HyperNetworks. *arXiv:1609.09106 [cs]*, December 2016.

Hochreiter, S., Younger, A., and Conwell, P. *Learning To Learn Using Gradient Descent*. September 2001. ISBN 978-3-540-42486-4. doi: 10.1007/3-540-44668-0_13.

Husken, M. and Goerick, C. Fast learning for problem classes using knowledge based network initialization. In *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium*, volume 6, pp. 619–624 vol.6, July 2000. doi: 10.1109/IJCNN.2000.859464.

Ioffe, S. and Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *arXiv:1502.03167 [cs]*, March 2015.

Kearns, M. J., Mansour, Y., and Ng, A. Y. Approximate Planning in Large POMDPs via Reusable Trajectories. pp. 7.

Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A. A., Milan, K., Quan, J., Ramalho, T., Grabska-Barwinska, A., Hassabis, D., Clopath, C., Kumaran, D., and Hadsell, R. Overcoming catastrophic forgetting in neural networks. *Proceedings of the National Academy of Sciences*, 114(13):3521–3526, March 2017. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1611835114.

Koch, G. Siamese Neural Networks for One-Shot Image Recognition. pp. 30.

Krähenbühl, P., Doersch, C., Donahue, J., and Darrell, T. Data-dependent Initializations of Convolutional Neural Networks. *arXiv:1511.06856 [cs]*, September 2016.

Langford, J. and Shawe-Taylor, J. PAC-Bayes & margins. In *Proceedings of the 15th International Conference on Neural Information Processing Systems, NIPS'02*, pp. 439–446, Cambridge, MA, USA, January 2002. MIT Press.

Li, K. and Malik, J. Learning to Optimize Neural Nets. *arXiv:1703.00441 [cs, math, stat]*, November 2017.

- 495 Maclaurin, D., Duvenaud, D., and Adams, R. P. Gradient-
 496 based Hyperparameter Optimization through Reversible
 497 Learning. pp. 10.
- 498 Majumdar, A., Farid, A., and Sonar, A. PAC-Bayes Con-
 499 trol: Learning Policies that Provably Generalize to Novel
 500 Environments. June 2018.
- 501 Maurer, A. A note on the pac bayesian theorem. *arXiv*
 502 preprint cs/0411099, 2004.
- 503 McAllester, D. A. Some pac-bayesian theorems. *Machine*
 504 *Learning*, 37(3):355–363, 1999a.
- 505 McAllester, D. A. Some PAC-Bayesian Theorems. *Machine*
 506 *Learning*, 37(3):355–363, December 1999b. ISSN 1573-
 507 0565. doi: 10.1023/A:1007618624809.
- 508 Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A.,
 509 Antonoglou, I., Wierstra, D., and Riedmiller, M.
 510 Playing Atari with Deep Reinforcement Learning.
arXiv:1312.5602 [cs], December 2013.
- 511 Naik, D. K. and Mammone, R. J. Meta-neural net-
 512 works that learn by learning. In *[Proceedings 1992]*
*513 IJCNN International Joint Conference on Neural Net-
 514 works*, volume 1, pp. 437–442 vol.1, June 1992. doi:
 515 10.1109/IJCNN.1992.287172.
- 516 Neu, G., Jonsson, A., and Gómez, V. A unified
 517 view of entropy-regularized Markov decision processes.
 518 *arXiv:1705.07798 [cs, stat]*, May 2017.
- 519 Neyshabur, B., Bhojanapalli, S., McAllester, D., and Sre-
 520 bro, N. Exploring Generalization in Deep Learning.
arXiv:1706.08947 [cs], July 2017.
- 521 Neyshabur, B., Bhojanapalli, S., and Srebro, N. A PAC-
 522 Bayesian Approach to Spectrally-Normalized Margin
 523 Bounds for Neural Networks. *arXiv:1707.09564 [cs]*,
 524 February 2018.
- 525 Nichol, A., Pfau, V., Hesse, C., Klimov, O., and Schulman, J.
 526 Gotta Learn Fast: A New Benchmark for Generalization
 527 in RL. *arXiv:1804.03720 [cs, stat]*, April 2018.
- 528 Pacelli, V. and Majumdar, A. Learning Task-Driven Control
 529 Policies via Information Bottlenecks. *arXiv:2002.01428*
 530 [*cs, math, stat*], February 2020.
- 531 Peng, X. B., Andrychowicz, M., Zaremba, W., and Abbeel,
 532 P. Sim-to-Real Transfer of Robotic Control with Dynam-
 533 ics Randomization. In *2018 IEEE International Confer-
 534 ence on Robotics and Automation (ICRA)*, pp. 3803–3810,
 535 May 2018. doi: 10.1109/ICRA.2018.8460528.
- 536 Ravi, S. and Larochelle, H. OPTIMIZATION AS A
 537 MODEL FOR FEW-SHOT LEARNING. pp. 11, 2017.
- 538 Reed, S., Chen, Y., Paine, T., van den Oord, A., Eslami,
 539 S. M. A., Rezende, D., Vinyals, O., and de Freitas, N.
 540 Few-shot Autoregressive Density Estimation: Towards
 541 Learning to Learn Distributions. *arXiv:1710.10304 [cs]*,
 542 February 2018.
- 543 Salimans, T. and Kingma, D. P. Weight Normalization:
 544 A Simple Reparameterization to Accelerate Training of
 545 Deep Neural Networks. *arXiv:1602.07868 [cs]*, June
 546 2016.
- 547 Saxe, A. M., McClelland, J. L., and Ganguli, S. Exact
 548 solutions to the nonlinear dynamics of learning in deep
 549 linear neural networks. *arXiv:1312.6120 [cond-mat, q-
 550 bio, stat]*, February 2014.
- 551 Schmidhuber, J. Learning to Control Fast-Weight Memories:
 552 An Alternative to Dynamic Recurrent Networks. *Neural*
553 Computation, 4(1):131–139, January 1992. ISSN 0899-
 554 7667. doi: 10.1162/neco.1992.4.1.131.
- 555 Schulman, J. Trust Region Policy Optimization. pp. 9.
- 556 Seeger, M. PAC-Bayesian Generalisation Error Bounds
 557 for Gaussian Process Classification. *Journal of Machine*
558 Learning Research, 3(Oct):233–269, 2002. ISSN ISSN
 559 1533-7928.
- 560 Sinha, A., Namkoong, H., Volpi, R., and Duchi, J. Cer-
 561 tifying Some Distributional Robustness with Principled
 562 Adversarial Training. *arXiv:1710.10571 [cs, stat]*, May
 563 2020.
- 564 Snell, J., Swersky, K., and Zemel, R. S. Prototypical Net-
 565 works for Few-shot Learning. *arXiv:1703.05175 [cs,*
566 stat], June 2017.
- 567 Song, X., Jiang, Y., Tu, S., Du, Y., and Neyshabur, B.
 568 Observational Overfitting in Reinforcement Learning.
arXiv:1912.02975 [cs, stat], December 2019.
- 569 Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I.,
 570 and Salakhutdinov, R. Dropout: A simple way to pre-
 571 vent neural networks from overfitting. *The Journal of*
572 Machine Learning Research, 15(1):1929–1958, January
 573 2014. ISSN 1532-4435.
- 574 Sun, Y., Yin, X., and Huang, F. Temple: Learning template
 575 of transitions for sample efficient multi-task rl, 2020.
- 576 Sutton, R. S., Barto, A. G., et al. *Introduction to reinforce-
 577 ment learning*, volume 135. MIT press Cambridge, 1998.
- 578 Tan, J., Zhang, T., Coumans, E., Iscen, A., Bai, Y.,
 579 Hafner, D., Bohez, S., and Vanhoucke, V. Sim-to-
 580 Real: Learning Agile Locomotion For Quadruped Robots.
arXiv:1804.10332 [cs], May 2018.

- 550 Thrun, S. and Pratt, L. Learning to Learn: Introduction and
551 Overview. In Thrun, S. and Pratt, L. (eds.), *Learning to*
552 *Learn*, pp. 3–17. Springer US, Boston, MA, 1998. ISBN
553 978-1-4615-5529-2. doi: 10.1007/978-1-4615-5529-2_1.
- 554
- 555 Tobin, J., Biewald, L., Duan, R., Andrychowicz, M., Handa,
556 A., Kumar, V., McGrew, B., Ray, A., Schneider, J., Welin-
557 der, P., Zaremba, W., and Abbeel, P. Domain Random-
558 ization and Generative Models for Robotic Grasping. In
559 *2018 IEEE/RSJ International Conference on Intelligent*
560 *Robots and Systems (IROS)*, pp. 3482–3489, October
561 2018. doi: 10.1109/IROS.2018.8593933.
- 562
- 563 Veer, S. and Majumdar, A. Probably Approximately Cor-
564 rect Vision-Based Planning using Motion Primitives.
565 *arXiv:2002.12852 [cs, eess, math]*, November 2020.
- 566
- 567 Vinyals, O., Blundell, C., Lillicrap, T., Kavukcuoglu, K.,
568 and Wierstra, D. Matching Networks for One Shot Learn-
569 ing. *Advances in Neural Information Processing Systems*,
570 29:3630–3638, 2016.
- 571
- 572 Wang, Z., Bapst, V., Heess, N., Mnih, V., Munos, R.,
573 Kavukcuoglu, K., and de Freitas, N. Sample Efficient
574 Actor-Critic with Experience Replay. *arXiv:1611.01224*
575 [*cs*], July 2017.
- 576
- 577
- 578
- 579
- 580
- 581
- 582
- 583
- 584
- 585
- 586
- 587
- 588
- 589
- 590
- 591
- 592
- 593
- 594
- 595
- 596
- 597
- 598
- 599
- 600
- 601
- 602
- 603
- 604

Integrating Independent and Centralized Multi-agent Reinforcement Learning for Traffic Signal Network Optimization

Zhi Zhang, Jiachen Yang, Hongyuan Zha

Georgia Institute of Technology

ABSTRACT

Traffic congestion in metropolitan areas is a world-wide problem that can be ameliorated by traffic lights that respond dynamically to real-time conditions. Recent studies that applied deep reinforcement learning (RL) to optimize single traffic lights have shown significant improvement over conventional control. However, optimization of global traffic flow over a large road network fundamentally is a cooperative multi-agent control problem. Centralized learning via single-agent RL is infeasible due to an exponential joint-action space, while independent learning suffers from environment non-stationarity. We propose QCOMBO, a simple yet effective multi-agent reinforcement learning (MARL) algorithm that combines the advantages of independent and centralized learning without their shortcomings. We ensure scalability by selecting actions from individually optimized utility functions, which are shaped to maximize global performance via a novel consistency regularization loss between individual utility and a global action-value function. Experiments on diverse road topologies and traffic flow conditions in the SUMO traffic simulator show competitive performance of QCOMBO versus recent state-of-the-art MARL algorithms. We further show that policies trained on small sub-networks can effectively generalize to larger networks under different traffic flow conditions, providing empirical evidence for the suitability of MARL for intelligent traffic control.

ACM Reference Format:

Zhi Zhang, Jiachen Yang, Hongyuan Zha Georgia Institute of Technology. 2020. Integrating Independent and Centralized Multi-agent Reinforcement Learning for Traffic Signal Network Optimization. In *Proc. of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2020)*, Auckland, New Zealand, May 9–13, 2020, IFAAMAS, 3 pages.

1 INTRODUCTION

With increasing urbanization, traffic congestion is a significant and costly problem [10, 15]. While early works proposed to optimize traffic light controllers based on expert knowledge and traditional model-based planning [4, 9, 18], there are promising recent results on applying flexible model-free methods in reinforcement learning (RL) [21] and deep RL, such as DQN in particular [16], to find optimal policies for traffic light controllers that dynamically respond to real-time traffic conditions [1, 7, 11, 24]. These works model a single traffic light as a Markov decision process (MDP) equipped with a discrete action space (e.g. signal phase change) and a continuous state space (e.g. vehicle waiting time, queue length), and

train a policy to optimize the expected return of an expert-designed reward function.

However, the single-agent RL perspective on traffic control optimization fails to account for the fundamental issue that optimizing global traffic flow over a densely connected road network is a cooperative multi-agent problem, where independently-learning agents face difficulty in finding global optimal solutions. Instead, all traffic light agents must act cooperatively to optimize the global traffic condition while optimizing their own individual reward based on local observations. On the other hand, existing work that adopt the multi-agent perspective on traffic signal optimization either fall back to independent learning [5, 12, 13] or resort to centralized optimization of coordinated agents [2, 23]. Independent learners [22] only optimize their own reward based on local observations, cannot optimize for global criteria (e.g., different priorities for different intersections), and they face a nonstationary environment due to other learning agents, which violates stationarity assumptions of RL algorithms. Therefore, these approaches do not account for the importance of macroscopic measures of traffic flow [8]. While centralized training can leverage global information, it requires maximization over a combinatorially-large joint action space and hence is difficult to scale. Motivated by these challenges, our paper focuses on deep multi-agent reinforcement learning (MARL) for traffic signal control with the following specific contributions:

1. Novel objective function combining independent and centralized training. We propose QCOMBO, a Q-learning based method with a new objective function that combines the benefits of both independent and centralized learning (Figure 1). We extended the definition of a single-agent reward [24] by defining the global reward as a weighted sum of individual rewards using the PageRank algorithm [17] to decide the weights. The key insight is to learn a global action-value function using the global reward, employ agent-specific observations and local rewards for fast independent learning of local utility functions, and enforce consistency between local and global functions via a novel regularizer. Global information shapes the learning of local utility functions that are used for efficient action selection.

2. Evaluation of state-of-the-art MARL algorithms on traffic signal optimization. Recent work proposed more sophisticated deep MARL algorithms for cooperative multi-agent problems with a global reward [6, 19, 20], under the paradigm of centralized training with decentralized execution [3]. However, as they were not designed for settings with individual rewards, it is open as to whether performance can be surpassed by leveraging agent-specific information. While they have shown promise on video game tasks, to the best of our knowledge they have not been tested on the important real-world problem of optimizing traffic signal over a

network. Hence we conducted extensive experiments comparing our algorithm versus independent Q-learning (IQL), independent actor-critic (IAC), COMA [6], VDN [20] and QMIX [19].

3. Generalizability of traffic light control policies. To the best of our knowledge, we conduct the first investigation on the generalizability and transferability of deep MARL policies for traffic signal control. Given improvements in sensor technology, measurements of traffic conditions can be increasingly accurate and real-world measurements can approach ideal simulated data. Hence, there is strong motivation to investigate whether a decentralized policy trained with simulated traffic approximating real-world conditions can be transferred to larger networks and different traffic conditions without loss of performance.

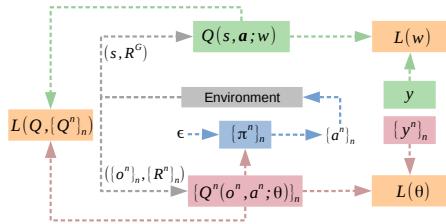


Figure 1: QCOMBO architecture combining independent learning of $Q^n(o^n, a^n)$ with centralized training of $Q(s, a)$ via a novel consistency loss $L(Q, \{Q^n\}_n)$

2 ARCHITECTURES FOR QCOMBO

QCOMBO is a novel combination of centralized and independent learning with coupling achieved via a new consistency regularizer. We optimize a composite objective (1) consisting of three parts: an individual term based on the loss function of independent DQN (2), a global term for learning a global action-value function (3), and a shaping term that minimizes the difference between the weighted sum of individual Q values and the global Q value (6), where λ controls the extent of regularization.

$$\mathcal{L}_{tot}(w, \theta) = \mathcal{L}(w) + \mathcal{L}(\theta) + \lambda \mathcal{L}_{reg} \quad (1)$$

$$\mathcal{L}(\theta^n) = \frac{1}{N} \sum_{n=1}^N \mathbb{E}_{\pi} \left[\frac{1}{2} (y_t^n - Q_{\theta^n}^n(o_t^n, a_t^n))^2 \right] \quad (2)$$

$$\mathcal{L}(w) = \mathbb{E}_{\pi} \left[\frac{1}{2} (y_t - Q_w^\pi(s_t, a_t))^2 \right] \quad (3)$$

$$Q^\pi(s, a) := \mathbb{E}_{\pi} \left[\sum_{t=0}^{\infty} \gamma^t R^g \mid s_0 = s, a_0 = a \right] \quad (4)$$

$$y_t = R_t^g + \gamma Q_w^\pi(s', a')|_{a'^m=\text{argmax}_{a^n} Q_\theta^n(o^n, a^n)} \quad (5)$$

$$\mathcal{L}_{reg} := \mathbb{E}_{\pi} \left[\frac{1}{2} (Q_w^\pi(s, a) - \sum_{n=1}^N k^n Q_\theta^n(o^n, a^n))^2 \right] \quad (6)$$

Q_w^π (4) and Q_θ^n are global and individual utility functions, y_t^n, y_t are the individual and global TD target.

By optimizing individual utility functions Q^n instead of a global optimal Q function, we reduce the maximization problem at each step of Q-learning from $O(|\mathcal{A}|^N)$ to $O(N|\mathcal{A}|)$. We also learn the global Q function under the joint policy induced by all agents' local utility functions, rather than learn the optimal global Q function, and use it to shape the learning of individual agents via information in global state s and global reward R^g . Crucially, action selection for

computing the TD target (5) uses the greedy action from local utility functions and does not use the global Q function. The collection of local utility functions induce a joint policy π that generates data for off-policy learning of the global action-value function Q^π . The regularization brings the weighted sum of individual utility functions closer to global expected return, so that the optimization of individual utility functions is influenced by the global objective rather than purely determined by local information.

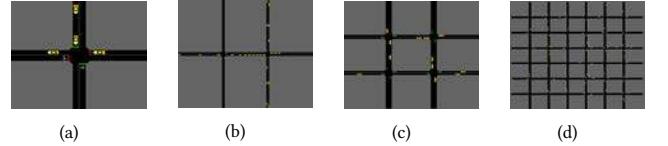


Figure 2: Grid topology used : (a) 1 traffic light example; (b) 2 traffic lights; (c) 2x2 traffic lights; (d) 6x6 traffic lights

3 EXPERIMENTAL SETUP

We formulate the multi-agent traffic light control problem as a partially-observed Markov game, consisting of N agents (Figure 2). Each agent controls the phase of one traffic light at an intersection.

We evaluated the performance of our method against a large set of baselines on multiple road networks under a variety of traffic conditions in the SUMO simulator [14, 25]. We implemented all algorithms using deep neural networks as function approximators. For each algorithm, we report the mean of five independent runs.

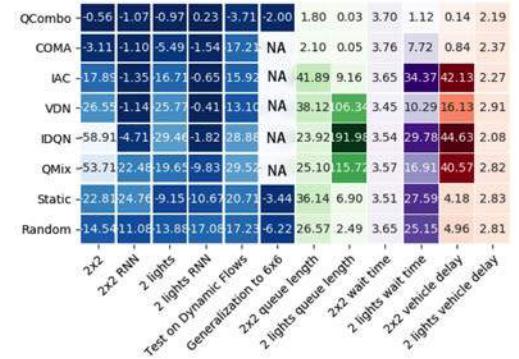


Figure 3: Heat map showing each algorithm's final performance, the left six columns are final reward under different road networks, the rest are measures of traffic conditions

4 RESULTS AND CONCLUSIONS

Over all flow and network configurations, QCOMBO attained the global optimal performance and is most stable among all algorithms (Figure 3). The performance of QCOMBO on test conditions does not heavily depend on specific choices of training conditions. Experiments also indicate that QCOMBO can be generalized with limited loss of performance to large traffic networks. Our work gives strong evidence for the feasibility of training cooperative policies for generalizable, scalable and intelligent traffic light control.

REFERENCES

- [1] Baher Abdulhai, Rob Pringle, and Grigoris J Karakoulas. 2003. Reinforcement learning for true adaptive traffic signal control. *Journal of Transportation Engineering* 129, 3 (2003), 278–285.
- [2] Itamar Arel, Cong Liu, T Urbanik, and AG Kohls. 2010. Reinforcement learning-based multi-agent system for network traffic signal control. *IET Intelligent Transport Systems* 4, 2 (2010), 128–135.
- [3] Daniel S Bernstein, Robert Givan, Neil Immerman, and Shlomo Zilberstein. 2002. The complexity of decentralized control of Markov decision processes. *Mathematics of operations research* 27, 4 (2002), 819–840.
- [4] Seung-Bae Cools, Carlos Gershenson, and Bart D’Hooghe. 2013. Self-organizing traffic lights: A realistic simulation. In *Advances in applied self-organizing systems*. Springer, 45–55.
- [5] Samah El-Tantawy, Baher Abdulhai, and Hossam Abdelgawad. 2013. Multiagent reinforcement learning for integrated network of adaptive traffic signal controllers (MARLIN-ATSC): methodology and large-scale application on downtown Toronto. *IEEE Transactions on Intelligent Transportation Systems* 14, 3 (2013), 1140–1150.
- [6] Jakob N Foerster, Gregory Farquhar, Triantafyllos Afouras, Nantas Nardelli, and Shimon Whiteson. 2018. Counterfactual multi-agent policy gradients. In *Thirty-Second AAAI Conference on Artificial Intelligence*.
- [7] Wade Genders and Saiedeh Razavi. 2016. Using a deep reinforcement learning agent for traffic signal control. *arXiv preprint arXiv:1611.01142* (2016).
- [8] Nikolas Geroliminis, Carlos F Daganzo, et al. 2007. Macroscopic modeling of traffic in cities. In *86th Annual Meeting of the Transportation Research Board, Washington, DC*.
- [9] Carlos Gershenson. 2004. Self-organizing traffic lights. *arXiv preprint nlin/0411066* (2004).
- [10] Federico Guerrini. 2014. Traffic Congestion Costs Americans \$124 Billion A Year, Report Says. *Forbes*, October 14 (2014).
- [11] Li Li, Yisheng Lv, and Fei-Yue Wang. 2016. Traffic signal timing via deep reinforcement learning. *IEEE/CAA Journal of Automatica Sinica* 3, 3 (2016), 247–254.
- [12] Mengqi Liu, Jiachuan Deng, Ming Xu, Xianbo Zhang, and Wei Wang. 2017. Cooperative Deep Reinforcement Learning for Traffic Signal Control. (2017).
- [13] Ying Liu, Lei Liu, and Wei-Peng Chen. 2017. Intelligent traffic light control using distributed multi-agent Q learning. *arXiv preprint arXiv:1711.10941* (2017).
- [14] Pablo Alvarez Lopez, Michael Behrisch, Laura Bicker-Walz, Jakob Erdmann, Yun-Pang Flötteröd, Robert Hilbrich, Leonhard Lücke, Johannes Rummel, Peter Wagner, and Eva-Maria Wiesßner. 2018. Microscopic Traffic Simulation using SUMO, In The 21st IEEE International Conference on Intelligent Transportation Systems. *IEEE Intelligent Transportation Systems Conference (ITSC)*. <https://elib.dlr.de/124092/>
- [15] Linsey McNew. 2014. Americans will waste \$2.8 trillion on traffic by 2030 if gridlock persists. (2014). <http://inrix.com/press-releases/americans-will-waste-2-8-trillion-on-traffic-by-2030-if-gridlock-persists/>
- [16] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. 2015. Human-level control through deep reinforcement learning. *Nature* 518, 7540 (2015), 529.
- [17] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. 1999. *The PageRank citation ranking: Bringing order to the web*. Technical Report. Stanford InfoLab.
- [18] Isaac Porche and Stéphane Lafontaine. 1999. Adaptive look-ahead optimization of traffic signals. *Journal of Intelligent Transportation System* 4, 3-4 (1999), 209–254.
- [19] Tabish Rashid, Mikayel Samvelyan, Christian Schroeder, Gregory Farquhar, Jakob Foerster, and Shimon Whiteson. 2018. QMIX: Monotonic Value Function Factorisation for Deep Multi-Agent Reinforcement Learning. In *Proceedings of the 35th International Conference on Machine Learning*, 4295–4304.
- [20] Peter Sunehag, Guy Lever, Audrunas Gruslys, Wojciech Marian Czarnecki, Vincius Zambaldi, Max Jaderberg, Marc Lanctot, Nicolas Sonnerat, Joel Z Leibo, Karl Tuyls, et al. 2017. Value-decomposition networks for cooperative multi-agent learning. *arXiv preprint arXiv:1706.05296* (2017).
- [21] Richard S Sutton and Andrew G Barto. 2018. *Reinforcement learning: An introduction*. MIT press.
- [22] Ming Tan. 1993. Multi-agent reinforcement learning: Independent vs. cooperative agents. In *Proceedings of the tenth international conference on machine learning*, 330–337.
- [23] Elise Van der Pol and Frans A Oliehoek. 2016. Coordinated deep reinforcement learners for traffic light control. *Proceedings of Learning, Inference and Control of Multi-Agent Systems (at NIPS 2016)* (2016).
- [24] Hua Wei, Guanjie Zheng, Huaxiu Yao, and Zhenhui Li. 2018. Intellilight: A reinforcement learning approach for intelligent traffic light control. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 2496–2505.
- [25] Cathy Wu, Aboudy Kreidieh, Kanaad Parvate, Eugene Vinitsky, and Alexandre M Bayen. 2017. Flow: Architecture and benchmarking for reinforcement learning in traffic control. *arXiv preprint arXiv:1710.05465* (2017).

Integrating independent and centralized multi-agent reinforcement learning for traffic signal network optimization

Paper #1244

ABSTRACT

Traffic congestion in metropolitan areas is a world-wide problem that can be ameliorated by traffic lights that respond dynamically to real-time conditions. Recent studies that applied deep reinforcement learning (RL) to optimize single traffic lights have shown significant improvement over conventional control. However, optimization of global traffic flow over a large road network fundamentally is a cooperative multi-agent control problem. Centralized learning via single-agent RL is infeasible due to an exponential joint-action space, while independent learning suffers from environment non-stationarity. We propose QCOMBO, a simple yet effective multi-agent reinforcement learning (MARL) algorithm that combines the advantages of independent and centralized learning without their shortcomings. We ensure scalability by selecting actions from individually optimized utility functions, which are shaped to maximize global performance via a novel consistency regularization loss between individual utility and a global action-value function. Experiments on diverse road topologies and traffic flow conditions in the SUMO traffic simulator show competitive performance of QCOMBO versus recent state-of-the-art MARL algorithms. We further show that policies trained on small sub-networks can effectively generalize to larger networks under different traffic flow conditions, providing empirical evidence for the suitability of MARL for intelligent traffic control.

KEYWORDS

Deep Reinforcement Learning; Multi-Agent Reinforcement Learning

1 INTRODUCTION

With increasing urbanization, traffic congestion is a significant and costly problem [13, 19]. While early works proposed to optimize traffic light controllers based on expert knowledge and traditional model-based planning [6, 12, 23], there are promising recent results on applying flexible model-free methods in reinforcement learning (RL) [26] and deep RL, such as DQN in particular [20], to find optimal policies for traffic light controllers that dynamically respond to real-time traffic conditions [1, 10, 14, 31]. These works model a single traffic light as a Markov decision process (MDP) equipped with a discrete action space (e.g. signal phase change) and a continuous state space (e.g. vehicle waiting time, queue length), and train a policy to optimize the expected return of an expert-designed reward function.

However, the single-agent RL perspective on traffic control optimization fails to account for the fundamental issue that optimizing

global traffic flow over a densely connected road network is a cooperative multi-agent problem, where independently-learning agents face difficulty in finding global optimal solutions. For example, if an intersection with low vehicle density in the North-South direction selfishly lets East-West traffic flow with little interruption to maximize its own performance, it will cause severe issues for any adjacent intersection that has heavy North-South traffic. Instead, all traffic light agents must act cooperatively to optimize the global traffic condition while optimizing their own individual reward based on local observations.

On the other hand, existing work that adopt the multi-agent perspective on traffic signal optimization either fall back to independent learning [16] or resort to centralized optimization of coordinated agents [2, 30]. Independent learners [29] only optimize their own reward based on local observations, cannot optimize for global criteria (e.g., different priorities for different intersections), and they face a nonstationary environment due to other learning agents, which violates stationarity assumptions of RL algorithms. While centralized training can leverage global information, it requires maximization over a combinatorially-large joint action space and hence is difficult to scale.

Motivated by these challenges, our paper focuses on deep multi-agent reinforcement learning (MARL) for traffic signal control with the following specific contributions:

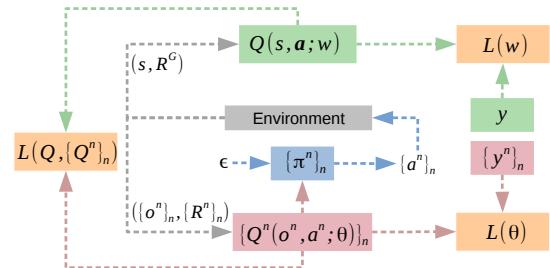


Figure 1: QCOMBO architecture combining independent learning of $Q^n(o^n, a^n)$ with centralized training of $Q(s, a)$ via a novel consistency loss $L(Q, \{Q^n\}_n)$

1. Novel objective function combining independent and centralized training. We propose QCOMBO, a Q-learning based method with a new objective function that combines the benefits of both independent and centralized learning (Figure 1). The key insight is to learn a global action-value function using the global reward, employ agent-specific observations and local rewards for fast independent learning of local utility functions, and enforce consistency between local and global functions via a novel regularizer. Global information shapes the learning of local utility functions that are used for efficient action selection.

2. Evaluation of state-of-the-art MARL algorithms on traffic signal optimization. Recent cooperative MARL algorithms specifically tackle the case when all agents share a single global reward [9, 24, 25]. However, as they were not designed for settings with individual rewards, it is open as to whether their performance can be surpassed by leveraging agent-specific information. While they have shown promise on video game tasks, to the best of our knowledge they have not been tested on the important real-world problem of optimizing traffic signal over a network. Hence we conducted extensive experiments comparing our algorithm versus independent Q-learning (IQL), independent actor-critic (IAC), COMA [9], VDN [25] and QMIX [24] on a variety of road networks with varying traffic conditions.

3. Generalizability of traffic light control policies. To the best of our knowledge, we conduct the first investigation on the generalizability and transferability of deep MARL policies for traffic signal control. Reinforcement learning methods are especially suitable for dynamic traffic light control since the transfer of a policy learned in simulation to real-world execution is arguably more feasible than in other domains (e.g. robotics). Similar to domains where deep RL excels [20], each traffic light has a small set of discrete actions for a signal phase change, which poses negligible issues for sim-to-real transfer. Given improvements in sensor technology, measurements of traffic conditions can be increasingly accurate and real-world measurements can approach ideal simulated data. Powerful model-free RL methods also do not require an accurate transition model that predicts traffic flow. Hence, there is strong motivation to investigate whether a decentralized policy trained with simulated traffic approximating real-world conditions can be transferred to larger networks and different traffic conditions without loss of performance.

2 RELATED WORK

Early work demonstrated the application of RL to single traffic light control [1]. The success of deep RL has spurred recent works that incorporate high dimensional state information into a more realistic problem definition [10, 15, 21, 31], which we further extend to define the observation and action spaces of our new multi-agent setting. Various choices of the reward function were proposed for training a single traffic light agent [3, 5, 21]. We extended the definition of a single-agent reward [31] by defining the global reward as a weighted sum of individual rewards using the PageRank algorithm [22].

Previous work on multi-agent traffic light control mostly relied on independent Q-learning (IQL) with heuristics to account for non-stationarity and coordination, such as: single-agent Q-learning for a central intersection surrounded by non-learning agents [2]; applying a Q function learned on a sub-problem to the full problem with max-plus action-selection [30]; training only one agent during each episode while fixing other agents' policies [16]; sharing information among neighboring Q-learning agents [7, 17]. These approaches do not account for the importance of macroscopic measures of traffic flow [11]. In contrast, our formulation explicitly shapes the learning of individual agents via a global reward.

Recent work proposed more sophisticated deep MARL algorithms for cooperative multi-agent problems with a global reward [9, 24, 25], under the paradigm of centralized training with decentralized execution [4]. However, these methods only learn from a global

reward without using available individual rewards, which motivates our proposal for a simple yet effective way to combine individual and centralized training. To the best of our knowledge, these algorithms have yet to be evaluated and compared in the real-world problem of multi-agent traffic light control, which we do as part of our main contributions.

3 MARKOV GAME FOR MULTI TRAFFIC LIGHT CONTROL

We formulate the multi-agent traffic light control problem as a partially-observed Markov game $\langle S, \{O\}^n, \{A\}^n, P, R, N, \gamma \rangle$, consisting of N agents labeled by $n = [1..N]$, defined as follows:

Agents $n \in [1..N]$. Each agent controls the phase of one traffic light at an intersection.

Observation space O^n . Since all traffic lights have the same measurement capabilities, all agents' observation spaces $O := O^1 = \dots = O^N$ have the same definition. Each agent's individual observation vector $o^n \in O$ depends on its own local traffic flow, with the following components: $q^n \in \mathbb{R}^l$, $v^n \in \mathbb{R}^l$, $wt^n \in \mathbb{R}^l$, $delay^n \in \mathbb{R}^l$ (for l incoming lanes at a traffic light), $ph^n \in \mathbb{R}^2$, and $d^n \in \mathbb{R}$, defined as:

- q^n : the length of queue on incoming lanes, defined as the total number of halting vehicles (speed less than 0.1m/s);
- v^n , the number of vehicles on each incoming lane;
- wt^n , the average waiting time of all vehicles on each incoming lane; defined as the time in minutes a vehicle spent with a speed below 0.1m/s since the last time it was faster than 0.1m/s
- $delay^n$, the average delay of all vehicles on each incoming lane, the delay of a lane is equal to 1 - (average vehicle speed)/(maximum allowed vehicle speed);
- ph^n : the traffic light's current phase, indicating the status of the east-west and north-south directions, represented by a one-hot variable $ph^n : EW \times NS \mapsto \{0, 1\}^2$;
- d^n : phase duration in seconds since the last phase change.

Prior work used an image representation of positions of all vehicles near a traffic light and required convolutional networks [31]. In contrast, we show this is not necessary and hence significantly reduce computational cost.

Global state space S contains all global information including every route and traffic light. Global state $s \in S$ is the concatenation of all local observation vectors.

Action space A^n . Traffic controllers with the same capabilities means $A := A^1 = \dots = A^N$. Extension to controllers with different types is easily done by learning separate Q functions for each type. The action a^n of each agent is a binary variable to indicate whether or not the traffic light will keep the current phase or switch to another phase. This definition is sufficient for policy learning because the agent's current phase is included in its own observation vector. The game has joint action space $A \equiv A^1 \times \dots \times A^n$. Agents produce joint action $a := (a^1, \dots, a^N) \in A$ at each time step. Let a^{-n} denote all actions *except* that of agent n .

Individual reward $R^n(s, a) : S \times A \mapsto \mathbb{R}$. We base our individual reward on previous work that used weighted route features as the reward for the single-agent traffic light setting [30, 31]. The individual reward $R^n = c_1 \times ql^n + c_2 \times wtl^n + c_3 \times dl^n + c_4 \times eml^n + c_5 \times fl^n + c_6 \times vl^n$ is a weighted linear combination of seven meaningful features that

capture the intuitive metrics of desirable and undesirable traffic conditions. The features are: $ql^n := \sum_{\Delta_t} \sum_l q_l^n$ is the sum of queue length of all incoming lanes, weighted by $c_1 = -0.5$, with $\Delta_t = 5s$. $wtl^n := \sum_{\Delta_t} \sum_l wtl_l^n$ is the sum of vehicle waiting time among all incoming lanes, with $c_2 = -0.5$. $dl^n := \sum_{\Delta_t} \sum_l delay_l^n$ is the sum of vehicle delay of all incoming lanes queue, with $c_3 = -0.5$. eml^n , # of vehicles that have emergence stops on the traffic light lanes over Δ_t , and $c_4 = -0.25$. fl^n , # of times that traffic lights change its phase over Δ_t , and $c_5 = -1$. vl^n , # of vehicles that pass the traffic light over Δ_t , and $c_6 = 1$.

Global reward R^g , defined as a weighted sum of individual rewards. We explored different methods to compute these weights. We used the PageRank algorithm to compute weights on each individual reward [22], since traffic intersections with higher risk of congestion are generally located in the central areas of the map that have higher interactions with surrounding traffic, and therefore should receive higher priority. Hence $R^g(s, a) := \sum_{n=1}^N k_n R^n(s, a)$, where $k_n = \text{PageRank}(n)$. While we considered using the traffic flow conditions under a fixed control policy to compute the weights for each traffic light, this is not a good choice since an arbitrary suboptimal nominal policy may produce a bad estimation of weights. In contrast, the PageRank algorithm accounts for the topological structure of the transportation network, addresses the connectivity and interaction between agents, and assigns higher weights to the rewards of highly-connected traffic lights.

Evaluation criteria. Given a reward function designed with sufficient expert domain knowledge and specified with enough precision to disambiguate different traffic states, we can investigate the performance of state-of-the-art MARL algorithms by directly evaluating them using the cumulative reward and components of the reward. Hence we do not resort to manual inspection of policy behavior, in contrast to previous work where certain states were aliased (i.e. produce the same reward) and manual inspection of the policy was required [31].

4 MARL ALGORITHMS

In this section, we give an overview of early and recent MARL algorithms, focusing on their respective strengths and weaknesses, with details in MARL algorithms of appendix 4. We use them as baselines for our experiments.

IQL and IAC. Independent Q-learning (IQL) and independent actor-critic (IAC) have demonstrated surprisingly strong performance in complex multi-agent systems [9, 24, 28, 29]. IQL directly applies single-agent Q-learning to each agent of the Markov game. While the optimal action-value function of an MDP is defined as

$$Q^*(s, a) := \max_{\pi} \mathbb{E}_{\pi} \left[\sum_{t=0}^{\infty} \gamma^t R_t \mid s_0 = s, a_0 = a \right], \quad (1)$$

IQL agents learn a local utility function $Q(o^n, a^n)$ by minimizing the loss function [20]

$$L(\theta) = \mathbb{E}_{\pi} \left[(R_t^n + \gamma \max_{a^n} Q_{\hat{\theta}}(o_{t+1}^n, a_t^n) - Q_{\theta}(o_t^n, a_t^n))^2 \right] \quad (2)$$

where θ are parameters of the function approximation and θ' are parameters of a target network. However, $Q(o^n, a^n)$ is not a true action-value function because the presence of other learning agents result in a nonstationary environment from any single agent's perspective.

Similarly, IAC directly applies the single-agent policy gradient with a variance-reduction baseline $b(s)$ [27]

$$\nabla_{\theta} J(\pi) = \mathbb{E}_{\pi} \left[\nabla_{\theta} \log \pi(a^n | o^n) (Q^{\pi}(o^n, a^n) - b(s)) \right] \quad (3)$$

to train an actor-critic pair for each agent, resulting in actors $\pi^n(a^n | o^n)$ and critics $Q^n(o^n, a^n)$. While IQL and IAC agents display strong ability to optimize individual rewards [29, 33], the lack of global information and a mechanism for cooperation means they are likely to settle for sub-optimal solutions.

COMA. In cooperative MARL with a single global reward, COMA [9] estimates a centralized action-value function $Q^{\pi}(s, a)$ to compute a counterfactual advantage function for a multi-agent policy gradient:

$$\nabla_{\theta} J(\pi) = \mathbb{E}_{\pi} \left[\sum_n \nabla_{\theta} \log \pi^n(a^n | o^n) (Q^{\pi}(s, a) - b(s, a^{-n})) \right] \quad (4)$$

$$b(s, a^{-n}) := \sum_{\hat{a}^n} \pi^n(\hat{a}^n | o^n) Q^{\pi}(s, (a^{-n}, \hat{a}^n)) \quad (5)$$

Their formulation is a low variance gradient estimate, as the advantage function evaluates the contribution of an agent's chosen action a^n versus the average of all possible counterfactuals \hat{a}^n , keeping other agents' a^{-n} fixed. However, since the only learning signal comes from the global reward and individual agents are not directly trained to improve local performance, COMA may exhibit slower training in cooperative traffic light control.

VDN. While IQL agents cannot learn to cooperate for a global reward, it is also not feasible to learn a single optimal action-value function $Q^*(s, a)$ since the maximization step requires searching over $|\mathcal{A}|^N$ joint actions. Instead, VDN [25] learns a joint action-value function that decomposes as $Q^{\text{VDN}}(s, a) := \sum_{n=1}^N Q^n(o^n, a^n)$, so that $Q^{\text{VDN}}(s, a)$ is trained with

$$L(\theta) = \mathbb{E}_{\pi} \left[(y_t - Q^{\text{VDN}}(s_t, a_t))^2 \right] \quad (6)$$

$$y_t := R + \gamma Q^{\text{VDN}}(s_{t+1}, a_{t+1})|_{a_{t+1}=\{\text{argmax}_{a^n} Q(o_{t+1}^n, a^n)\}_n} \quad (7)$$

Agents act greedily with respect to their own utility functions, while global reward is used for overall training. However, there is no guarantee in general settings that the true optimal $Q^*(s, a)$ can be decomposed as a linear combination of individually-optimized utilities, which could limit VDN's performance.

QMIX. QMIX [24] generalizes VDN by representing the optimal action-value function as a nonlinear function $Q^*(s, a) = F(Q^1, \dots, Q^N)$ of individual utility functions, while ensuring that the combination of individual argmax on each Q^n yields the same joint action as a global argmax on $Q^*(s, a)$. This is achieved by enforcing positive weights in the nonlinear mixing network F . Despite being a more expressive model than VDN, the stability of QMIX depends on appropriate choices of mixing network architecture, for which there is little theoretical guidance, and QMIX also relies on global reward without using local reward for training.

5 METHOD

We propose QCOMBO, a novel combination of centralized and independent learning with coupling achieved via a new consistency regularizer. We optimize a composite objective consisting of three parts: an individual term based on the loss function of independent DQN, a global term for learning a global action-value function, and

a shaping term that minimizes the difference between the weighted sum of individual Q values and the global Q value. This algorithm ensures that agents cooperate to maximize the global reward, which is difficult for independent learning agents to achieve, and also maintain the ability to optimize their individual performance using agent-specific observations and rewards, which is more efficient than a purely centralized approach.

5.1 Individual Part

Using individual observations and rewards for each agent is computationally efficient, since local observations generally have lower dimension than global state information, and also algorithmically efficient, since it avoids the difficulty of decomposing a single global reward signal for multi-agent credit assignment. By optimizing individual utility functions Q^n instead of a global optimal Q function, we reduce the maximization problem *at each step* of Q-learning from $O(|\mathcal{A}|^N)$ to $O(N|\mathcal{A}|)$. Parameterizing the local utility function for agent n with parameter θ^n , we minimize the loss

$$\mathcal{L}(\theta^n) = \frac{1}{N} \sum_{n=1}^N \mathbb{E}_{\pi} \left[\frac{1}{2} (y_t^n - Q_{\theta^n}^n(o_t^n, a_t^n))^2 \right] \quad (8)$$

$$y_t^n = R_t^n + \gamma \max_{\hat{a}^n} Q_{\hat{\theta}^n}^n(o_t^n, \hat{a}^n) \quad (9)$$

Since the agent population is homogeneous (i.e. all agents have the same observation and action spaces), we improve memory and computational efficiency by employing parameter-sharing among all agents, meaning that $\theta := \theta^n, \forall n \in [1..N]$ in all the subsequent text. Agents behave differently since they receive different observations, and we use an agent indicator as input for disambiguation. $\hat{\theta}$ are parameters of a target network [20].

5.2 Global Part

Without global information, independently learning agents face a nonstationary environment due to the presence of other learning agents, and they may have insufficient information to find cooperative optima. On the other hand, training an optimal global Q function is not scalable, since the Q-learning step would require maximization over $|\mathcal{A}|^N$ possible joint actions for N agents. To address this dilemma, our key insight is that we can learn the global Q function *under the joint policy induced by all agents' local utility functions*, rather than learn the *optimal* global Q function, and use it to shape the learning of individual agents via information in global state s and global reward R^g . Specifically, the joint policy defined by $\mathbf{a} \sim \pi(\mathbf{a}|s) = \{\text{argmax}_{a^n} Q^n(o^n, a^n)\}_{n=1}^N$ is associated with a global action-value function (letting $R_t^g := R_t^g(s, \mathbf{a})$):

$$Q^{\pi}(s, \mathbf{a}) := \mathbb{E}_{\pi} \left[\sum_{t=0}^{\infty} \gamma^t R^g \mid s_0 = s, \mathbf{a}_0 = \mathbf{a} \right] \quad (10)$$

Parameterizing $Q_w^{\pi}(s, \mathbf{a})$ with w , we minimize the loss:

$$\mathcal{L}(w) = \mathbb{E}_{\pi} \left[\frac{1}{2} (y_t - Q_w^{\pi}(s_t, \mathbf{a}_t))^2 \right] \quad (11)$$

$$y_t = R_t^g + \gamma Q_w^{\pi}(s', \mathbf{a}')|_{a'^n=\text{argmax}_{a^n} Q_{\theta}^n(o'^n, a^n)} \quad (12)$$

where we let $(\cdot)' := (\cdot)_{t+1}$. Crucially, action selection for computing the TD target (12) uses the greedy action from local utility functions and does not use the global Q function. The collection of local utility

functions induce a joint policy π that generates data for off-policy learning of the global action-value function Q^{π} . w are target network parameters.

5.3 Combined objective

If each agent greedily optimizes its own local utility function, the global return can be suboptimal. For example, if agent n (with low weight k^n) has no flow in the N-S direction while adjacent agent m (with high weight k^m) has heavy flow in the N-S direction, the individual optimal policy for n is to let W-E traffic flow continuously to m , which negatively impacts conditions at m and leads to low global reward. This is supported by experimental results in a 1x2 network. To address the suboptimality of independent learning, we propose a new consistency regularization loss

$$\mathcal{L}_{reg} := \mathbb{E}_{\pi} \left[\frac{1}{2} (Q_w^{\pi}(s, \mathbf{a}) - \sum_{n=1}^N k^n Q_{\theta}^n(o^n, a^n))^2 \right] \quad (13)$$

between global Q_w^{π} and individual utility functions Q_{θ}^n . Since Q_w^{π} is the true global action-value function with respect to the induced joint policy, this regularization brings the weighted sum of individual utility functions closer to global expected return, so that the optimization of individual utility functions is influenced by the global objective rather than purely determined by local information. Hence the regularizer prevents any individual agent from attaining high individual performance at the cost of collective performance.

The complete QCOMBO architecture (Figure 1) combines the individual loss (8), global loss (11), and consistency regularizer (13) into the overall objective:

$$\mathcal{L}_{tot}(w, \theta) = \mathcal{L}(w) + \mathcal{L}(\theta) + \lambda \mathcal{L}_{reg} \quad (14)$$

where λ controls the extent of regularization. Whenever any agent learns to attain high individual reward at the cost of global performance, which is likely when minimizing the individual loss, the consistency loss will increase to reflect inconsistency between individual and global performance; it will then decrease once global information influences the learning of individual Q^n . Our experiments provide evidence of this dynamic learning process that balances individual and global learning (Figure 2(a)). Since the third term is a regularizer, which in general is not necessarily zero at convergence, (14) does not force Q_w^{π} to equal the weighted sum of all Q_{θ} . Even at equality, agents still retain cooperation and are not independent because Q_w^{π} is trained using the total reward and (13) weighs each agent by k^n . Optimizing $\mathcal{L}(w)$ by itself does not enable action selection due to combinatorial explosion of the joint action space; optimizing $\mathcal{L}(\theta)$ alone amounts to IQL; and, crucially, removing our novel regularization term from (14) would decouple the global and individual losses and reduce (14) to IQL. Using global information to affect the learning of individual optimal policies is achieved when the two terms in \mathcal{L}_{reg} are close at convergence. However, this will not happen in general unless we include \mathcal{L}_{reg} in the overall loss function \mathcal{L}_{tot} in (14). This is precisely because the global Q_w^{π} is a function of the global state s and joint action a^1, \dots, a^n , while individual Q_{θ}^n is only a function of partial local observation o^n and individual action a^n . Under general nonlinear function approximation for both, and without imposing additional assumptions, Q_w^{π} is not equal to the weighted sum of Q_{θ}^n . This holds despite the fact that

$R^g = \sum_{n=1}^N k_n R^n$. Hence, we need the interaction term \mathcal{L}_{reg} to force the two terms to be close at convergence, which is experimentally confirmed by Figure 2(a).

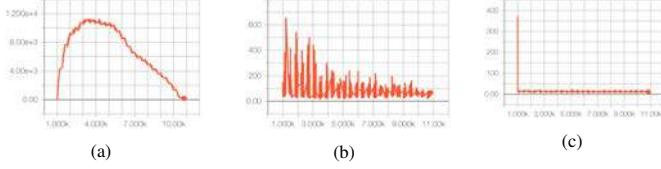


Figure 2: Learning curve for loss in QCOMBO (1×2 map): (a) consistency loss; (b) global loss; (c) individual loss

At each training step, we interleave the updates to θ and w to exchange information between the global and individual parts, allowing each agent to learn a policy that considers the effects of other agents' learning. The parameters w and θ are updated by gradient descent:

$$\begin{aligned} \nabla_w \mathcal{L}_{tot} &= E_{\pi} \left[-(y_t - Q_w^{\pi}(s_t, a_t)) \nabla_w Q_w^{\pi}(s_t, a_t) \right. \\ &\quad \left. + \lambda (Q_w^{\pi}(s_t, a_t) - \sum_n k^n Q_{\theta}^n(o_t^n, a_t^n)) \nabla_w Q_w^{\pi}(s_t, a_t) \right] \\ &= -E_{\pi} \left[[Y - (1 + \lambda) Q_w^{\pi}(s_t, a_t) \right. \\ &\quad \left. + \lambda \sum_n k^n Q_{\theta}^n(o_t^n, a_t^n)] \nabla_w Q_w^{\pi}(s_t, a_t) \right] \\ Y &:= R^g + \gamma Q_w^{\pi}(s', a')|_{a'=\{\text{argmax}_a Q_{\theta}^n(o'^n, a^n)\}} \end{aligned} \quad (15)$$

$$\begin{aligned} \nabla_{\theta} \mathcal{L}_{tot} &= -\mathbb{E}_{\pi} \left[\frac{1}{N} \sum_{n=1}^N (y_t^n - Q_{\theta}^n(o_t^n, a_t^n)) \nabla_{\theta} Q_{\theta}^n(o_t^n, a_t^n) \right. \\ &\quad \left. + \lambda (Q_w^{\pi}(s_t, a_t) - \sum_{m=1}^N k^m Q_{\theta}^m(o_t^m, a_t^m)) \nabla_{\theta} \sum_{n=1}^N k^n Q_{\theta}^n(o_t^n, a_t^n) \right] \\ &= -\mathbb{E}_{\pi} \left[\sum_{n=1}^N (Z_1 + \lambda k^n Z_2) \nabla_{\theta} Q_{\theta}^n(o_t^n, a_t^n) \right] \\ Z_1 &:= \frac{1}{N} \left(R_t^n + \gamma \max_{\hat{a}^n} Q_{\theta}^n(o'^n, \hat{a}^n) - Q_{\theta}^n(o_t^n, a_t^n) \right) \\ Z_2 &:= Q_w^{\pi}(s_t, a_t) - \sum_{m=1}^N k^m Q_{\theta}^m(o_t^m, a_t^m) \end{aligned} \quad (16)$$

6 EXPERIMENTAL SETUP

We evaluated the performance of our method against a large set of baselines (described in Section 4) on multiple road networks under a variety of traffic conditions in the SUMO simulator [18, 32]. We describe all key experimental setup details in this section. Section 7 provides detailed analysis of each algorithm's performance. For each algorithm, we report the mean of five independent runs, with standard deviation reported in Table 2.

Table 1: Network topologies, flow schedules and flow rates in all training and testing configurations

Program And Topology	Time Period A:train B:test	Num Vehicles/Hour	
		Horizontal (Bot to Top)	Vertical (Left to Right)
1 × 2 two Agents	A:0-12000	700	10, 620
2 × 2 Same State	A:0-12000	700, 700	700, 700
Generalization different flow 1 2 × 2	A:0-12000 B:0-2000 B:2000-4000 B:4000-6000	700, 280 700, 280 1000, 800 1400, 1000	10, 620 10, 620 900, 700 400, 900
Generalization different flow 2 2 × 2	A:0-12000 B:0-2000 B:2000-4000 B:4000-6000	1000, 580 1000, 580 1000, 800 1400, 1000	110, 920 110, 920 900, 700 400, 900
Generalization different flow 3 2 × 2	A:0-12000 B:0-2000 B:2000-4000 B:4000-6000	700, 700 700, 700 1000, 800 1400, 1000	700, 700 700, 700 900, 700 400, 900
Train on 2 × 2 test on 6 × 6	A:0-12000 B:0-12000	700, 700 700, 280, 260, 240, 780, 200	700, 700 10, 620, 620, 50, 90, 700

6.1 Environment

We used the Flow framework [32] with the SUMO simulator. We used homogeneous vehicles of the same type and size in all experiments. Extension to heterogeneous vehicles requires no modification to our algorithm, but only redefinition of observation vectors (e.g., a longer vehicle takes up two units in queue length). Road networks are defined as the intersection of m horizontal and n vertical roads (e.g. a 1×2 network has one horizontal route intersecting two vertical routes.). Each learning agent is a traffic light at an intersection. Each road between two intersections is 400m long and has two lanes with opposite directions of travel. Hence each agent has one incoming and one outgoing lane for each edge. Vehicles are emitted at the global outer boundaries of each edge with random starting lane and fixed entering speed. We used many traffic flows that differ in the number of vehicles per hour in different time periods, described in Table 1 for all traffic configurations. At each time step, the traffic light is green exclusively for either the horizontal or the vertical direction.

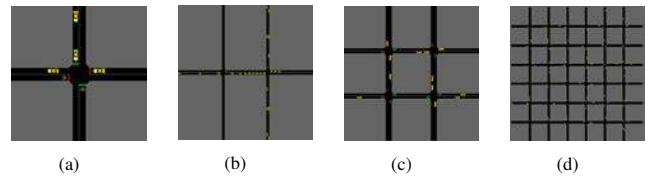


Figure 3: Grid topology used : (a) 1 traffic light example; (b) 2 traffic lights; (c) 2 × 2 traffic lights; (d) 6 × 6 traffic lights

Uniform flow on symmetric 2×2 network. The first road network is a symmetric environment with $N = 4$ traffic light agents

defined by a 2×2 network Figure 3(c). Vehicle emissions from all route boundaries are equal and uniform throughout the entire training horizon. Due to symmetry, all agents' rewards are weighted equally, and all agents should learn the same policy or value functions.

Non-uniform traffic flow on 1×2 network. The second traffic control scenario has more challenging traffic dynamics. One horizontal route and two vertical routes form two interactions (Figure 3(b)). The horizontal route has 700 vehicles/hour, one vertical route on the left has only 10 vehicles/hour, while the second vertical route on the right has 600 per hour. This is a common real-world scenario where one main arterial road with dense traffic is adjacent and parallel to a smaller local road with sparse traffic. The greedy strategy for the left traffic light is to let vehicles in the E-W direction pass almost all the time, which would lead to high incoming E-W traffic for the second traffic light that already has heavy N-S traffic. Hence cooperative learning is necessary for the left light to close the E-W route to some extent, to optimize global performance.

Generalization To Different Flows. We investigated how well policies learned by each algorithm in one traffic condition generalize to different traffic conditions without further training. This is crucial for real-world applicability since training on every possible traffic condition is not feasible. In the first experiment, we trained a policy under a *static* traffic flow in the 2×2 network, but tested it on three consecutive equal-duration time periods with *different vehicle densities*. Denoting the flow as a vector $f := (\text{bot}, \text{up}, \text{left}, \text{right}) \in \mathbb{R}^4$ specifying the number of vehicles per hour on each vertical and horizontal route, traffic flows in the second and third test periods are $f_{t_2} := (1000, 800, 900, 700)$, $f_{t_3} := (1400, 1000, 400, 900)$ (Table 1 third row). The first test period has the same traffic flow as the training phase.

We further investigated the extent to which generalization performance is affected by the specific traffic condition used in training. Specifically, we trained separate QCOMBO policies with *different flows* in the same network topology (with one policy per training flow), and test all policies under the *same* flow. We denote the i -th train-test combination ($i = 1, 2, 3$) as 4-component sequence $F^i = \{f_{t_0}^i, f_{t_1}^i, f_{t_2}^i, f_{t_3}^i\}$, where the first flow is the training flow followed by three test flows, and f_{t_2} and f_{t_3} are shared by all train-test combinations. Then the three train-test programs are: $F^1 = \{f_{t_0}^1, f_{t_1}^1, f_{t_2}, f_{t_3}\}$, $F^2 = \{f_{t_0}^2, f_{t_1}^2, f_{t_2}, f_{t_3}\}$, $F^3 = \{f_{t_0}^3, f_{t_1}^3, f_{t_2}, f_{t_3}\}$, where $f_{t_0}^1 = f_{t_1}^1 := (700, 280, 10, 620)$, $f_{t_0}^2 = f_{t_1}^2 := (1000, 580, 110, 920)$, $f_{t_0}^3 = f_{t_1}^3 := (700, 700, 700, 700)$ are training flows and the testing flow in the first 1/3 period. Three programs use the same testing flow for the second and third periods (f_{t_2}, f_{t_3}), specified in Table 1. Time periods t_0, t_1, t_2, t_3 are 10000, 1000, 1000, 1000 steps, respectively.

Generalization to larger networks. Directly training on simulations of large real-world traffic networks may not be computationally practical due to the combinatorially-large state and joint-action spaces, regardless of independent or centralized training. Previous approaches first formulated simple models for regional traffic and then realigned to the whole system [8], or relied on transfer planning [30]. In contrast, we investigated the feasibility of training on a sub-network and directly transferring the learned policies without further training into a larger network. Direct deployment in a larger systems is possible as QCOMBO learns decentralized policies. We constructed a 6×6 traffic network with 36 traffic lights (Figure 3(d))

and nonuniform traffic flows, which severely reduces the possibility that any traditional hand-designed traffic control plan can be the optimal policy. Policies trained via QCOMBO in the 2×2 network were directly tested on the 6×6 network.

6.2 Algorithm implementations

Neural Network Architecture. We implemented all algorithms using deep neural networks as function approximators. We ensure that all policy, value, and action-value functions have the same neural network architecture among all algorithms—to the extent allowed by each algorithm (e.g. hypernetwork architecture required by QMIX)—for fair comparison. The individual utility functions Q^n of IDQN, VDN, QMIX, and QCOMBO are represented by fully-connected three-layer neural networks, with 256 units in the hidden layers and ReLu activation, where the last layer has $|A|$ output nodes. QMIX has a two level hypernetwork that generates the weight matrix and bias vector from the global state s , to compute the inner product with each Q^n and produce one state action value $Q(s, a)$. VDN takes the sum of Q^n to calculate the total Q , which is minimized by the squared error loss. The critic of IAC is a value function V , which is used to estimate the TD error. The COMA uses a centralized Q minus a counterfactual baseline to compute the COMA policy gradient. The critics of COMA and IAC have the same three-layer neural network structure, similar to the Q-functions in IDQN, VDN, QMIX. Both IAC and COMA use the same actor network to approximate $\pi(a_t | o_t)$, which is also a three-layer fully connected neural network with a softmax activation in the last layer. We give the agent's observations, last actions, and one-hot vector of agent labels as input to utility functions (Q^n of QCOMBO, IDQN, QMIX, and VDN, V^n of IAC). Q^π is approximated by a three-layer fully-connected neural network with 256 hidden units. We give the global state s as the inputs of QCOMBO, and we give s , all other agents' actions a^{-n} as inputs of COMA. Motivated by the possibility of periodic behavior of traffic, we also experimented with RNN and GRU cells for Q^n of IDQN, VDN, QMIX and QCOMBO, and the $\pi(a_t | o_t)$ of IAC and COMA. We set up RNN's GRU cell having 64 hidden units with a Relu activation and let the hidden state retained for each training step and re-used as input for the next training batch. For the updating of the value function networks, the target network parameters are gradually updated by taking 1% of online network and 99% of current target network.

Training Strategy. All the algorithms have same training strategy. The total time horizon for training and online evaluation is limited to 12K SUMO time steps. We limit the agents to make decisions for every 5 times steps to avoid high-frequency flickering of traffic lights. For each training workflow, the environment runs freely for 1000 time steps using a random policy to populate the road network with vehicles before training begins. Then we alternate between training and evaluation, with each training and evaluation cycle lasting 200 time steps. In every training step (which equals to 5 SUMO time steps, so there are 40 training steps per cycle), the algorithm randomly collects 100 minibatches and each minibatch contains 30 samples, from the replay buffer, and for each minibatch it updates parameters of the the current networks via stochastic gradient descent using a 0.001 learning rate. The Adam optimizer is used for optimization. The replay buffer is a first-in first-out queue

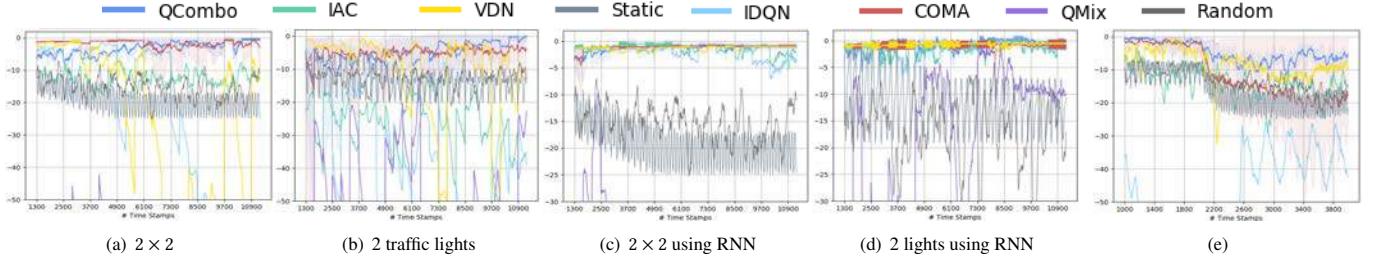


Figure 4: (a)-(d) show rewards among algorithms ;(e) shows generalization to different traffic flows

containing the 200 most recent samples. If the RNN framework is involved, we use a different sampling strategy: 1. We let every training cycle contains 6 consecutive periods; 2. in each period, we collect 30 consecutive samples and feed them into the RNN in the order of their corresponding time steps from the earliest to the latest; 3. after we finish the parameter updating in one period, we keep the hidden state and use it to initialize the hidden state for the next period. The current action is selected through an ϵ -greedy behavior policy with ϵ starting with 0.9 and decayed by a geometric factor of 0.995 between two training steps and fixed within one training step. A training cycle is followed with an on-policy evaluation cycle, where we let the algorithm execute the current policy and stop decaying the ϵ and we record the rewards and other information for evaluation of the policy.

We provide our code and configuration file at: download link, for interested readers' reference.

7 RESULTS

To analyze differences in algorithms, we show full learning curves for all algorithms in addition to reporting final performance. This is crucial because previous work on deep RL for traffic signal control noted the potential for instability during training [30]. Our learning curves were generated by conducting evaluation measurements periodically during training (i.e. test the current policy for 400 steps). Since it takes time to populate the road networks, our learning curves start after 1000 simulation time steps. For every experiment, we ran a static policy (change light phase every 30s), and a random policy (keep or change the current phase every 5s) so that improvements due to learning can be clearly seen.

7.1 Static Traffic Flows

Figure 4 shows the global reward in both the 2×2 and 1×2 road networks using both fully-connected and RNN neural networks. The shaded regions in Figure 4 show the 95% confidence interval of QCOMBO and COMA, computed across seeds. We omitted other algorithms' confidence intervals to reduce clutter and make the figures easier to inspect. Over all flow and network configurations, QCOMBO attained the global optimal performance and is most stable among all algorithms. Policy-based methods COMA and IAC, which mitigate issues with state aliasing and are more robust to small changes of observation, show lower variance and higher average reward than other value-based methods (IDQN, VDN and QMIX).

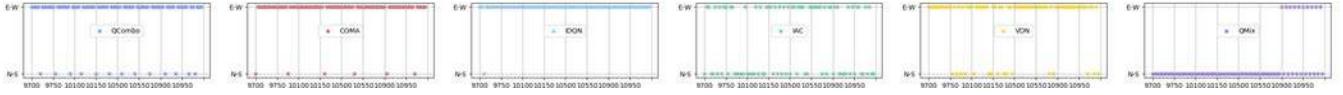
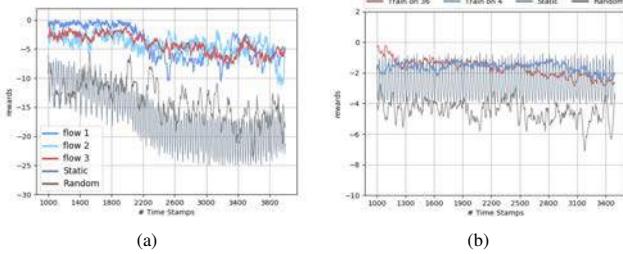
In the 2×2 environment Figure 4(a), QCOMBO and COMA converged to global optimal policies. Both two maintain good traffic condition throughout learning, as reflected by queue length, waiting

Table 2: Final Traffic Condition After Learning

	Queue Length	Wait Time	Vehicle Delay
2×2 balanced			
QCOMBO	1.8(0.1)	0.0(0.0)	3.3(0.0)
IDQN	23.9(24.9)	192.0(270.6)	3.4(0.3)
IAC	41.9(1.7)	9.2 (1.3)	3.4(0.0)
COMA	2.1(0.3)	0.1(0.0)	3.3(0.0)
QMIX	25.1(32.1)	115.7(163.6)	3.6(0.3)
VDN	38.1(29.1)	106.3(105.8)	3.4(0.3)
Random	26.6(0.0)	2.5(0.0)	3.6(0.0)
Static	36.1(0.0)	6.9(0.0)	3.5(0.0)
1×2 unbalanced			
QCOMBO	1.1(0.9)	0.1(0.2)	2.1(0.1)
IDQN	29.8(22.9)	44.6 (52.6)	2.4(0.9)
IAC	34.4(33.2)	42.1(78.9)	2.3(1.1)
COMA	7.7(8.5)	0.8(1.0)	2.2(1.2)
QMIX	16.9(19.1)	40.6(58.3)	2.8(0.1)
VDN	10.3(15.1)	16.1(34.1)	2.9(0.4)
Random	25.1(7.7)	4.9(1.4)	2.8(0.5)
Static	27.6(4.5)	4.2(1.6)	2.8(0.0)

time, and delay time of vehicles in Table 2. VDN performed better than the random and static policies, but worse than QCOMBO and COMA, as it can only learn a restricted set of linearly-combined Q functions. IDQN and IAC were not stable and failed to reach a global optimum, showing that learning cooperation from the global reward is necessary. QMIX diverged possibly due to the difficulty of stabilizing its hypernetwork. All algorithms (with the exception of QMIX) improved with the use of RNN for policy or value functions (Figure 4(c)), giving strong evidence that RNNs are especially suitable for handling history information and periodicity of traffic.

Performance differences between algorithms are more apparent in the 1×2 environment with non-uniform traffic flow (Figure 4(b)). QCOMBO converged to the optimal policy, exceeding the performance of all other algorithms. VDN began with high performance but struggled to maintain a good policy as more vehicles enter the road. Our comparison to IDQN acts as an ablation study that demonstrates the effectiveness of the regularization loss. QCOMBO's higher performance and stability over IDQN shows that the new regularization loss in QCOMBO helps to stabilize learning of independent utility functions by limiting their deviation from the centralized action-value function. As Section 5.3 explains, removing the regularization term \mathcal{L}_{reg} would decouple the global and individual loss terms, such that global information no longer affects the training

Figure 5: Traffic light phase selected by the left agent in the 1×2 topologyFigure 6: More Generalization Results, (a) Impact of training conditions on performance in new test condition; (b) Policy trained in 2×2 network generalizes to 6×6 network

of individual utility functions Q_θ^n . Since agents choose actions via greedy policies induced by Q_θ^n , removing \mathcal{L}_{reg} reduces our method to IDQN. The benefit of centralized learning (e.g. QCOMBO and COMA) over independent learning (e.g. IDQN, IDQN) is more apparent than in the 2×2 environment, since the non-uniform flow increases the impact of each agent on other agents' performance, resulting in the need for cooperation. IDQN, VDN and IAC exhibits oscillation, similar to behavior reported in [30]. QMIX suffers from convergence issues when using a fully-connected network (RNN results shown in Figure 4(d)).

We further explain differences in algorithm performance in 1×2 by analyzing actions (E-W or N-S phase) selected by the learned policies. In order to achieve cooperation, the left traffic light should open N-S and close E-W periodically to reduce incoming E-W traffic for the right light, who already experiences heavy N-S traffic. Figure 5 shows that QCOMBO and COMA achieve cooperation by turning off E-W traffic periodically (with low frequency, since it receives higher E-W than N-S traffic). However, IDQN greedily keeps E-W open for long durations, which is not globally optimal; IAC switches between the two phases almost equally; VDN switches to N-S too often than necessary; QMIX incorrectly chooses N-S more frequently than E-S.

7.2 Generalizing To Dynamic Traffic Flows

QCOMBO displayed the highest generalization performance, when trained on one traffic condition and deployed on two different test conditions Figure 4(e). We used the policy learned without RNN and under different seeds. As reflected by the decrease in performance of all policies when traffic flow changes at time step 2000, the test conditions were more difficult. While COMA and QCOMBO have equal performance on the training flow (first 1000 steps), QCOMBO generalized much more gracefully to the test conditions as the consistency regularizer prevents overfitting to local training conditions. VDN perform worse on the training flow than COMA but generalized better, despite experiencing a large drop in the second test flow when the vehicle density increases. IAC shows high variance on the

test condition, while IDQN and QMIX could not adapt to new flows due to low training performance.

Figure 6(a) shows results of the second generalization experiment, where we evaluated QCOMBO policies trained on three different traffic flows, using the same test program. We did not include other algorithms since the previous experiments demonstrate QCOMBO performs better than or equal to all others. The first 1000 steps have the same flow as during training, while the new flows appear at $t = 2000$ and $t = 3000$. QCOMBO policies show flow invariance during the t_2 and t_3 period: trained on different traffic conditions, generalization performance on new unseen conditions exhibit only small variability. This gives evidence that performance of QCOMBO on test conditions does not heavily depend on specific choices of training conditions.

7.3 Generalization To Larger Traffic Topology

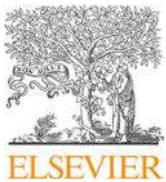
We directly applied the QCOMBO policy trained in the 2×2 traffic network without RNN to the 6×6 network with 36 traffic light agents, which poses a significant generalization challenge due to the decreased observability for any particular agent and the different traffic flow induced by the different network topology. Figure 6(b) shows that QCOMBO's policy is able to maintain high and stable test performance, with almost negligible difference from its training performance. Surprisingly, it sometimes attains higher reward even than a policy that was trained specifically on the 6×6 environment. This shows that centralized training with few agents can still produce policies that generalize well to larger settings, mitigating concerns about scalability of centralized training. This is strong evidence that a policy trained on a subset of a city road network can be deployed with little loss of performance at a larger scale.

7.4 Conclusion

We proposed QCOMBO, a novel MARL algorithm for traffic light control. QCOMBO combines the benefits of independent and centralized training, with a novel objective function that enforces consistency between a global action-value function and the weighted sum of individual optimal utility functions. We conducted detailed empirical evaluation of state-of-the-art MARL algorithms (IDQN, IAC, VDN, COMA, and QMIX) on network traffic light control under different map topologies and traffic flows, and showed that QCOMBO is a simple yet competitive approach to this real-world problem. Experiments also indicate that QCOMBO can be generalized with limited loss of performance to large traffic networks. Our work gives strong evidence for the feasibility of training cooperative policies for generalizable, scalable and intelligent traffic light control.

REFERENCES

- [1] Baher Abdulhai, Rob Pringle, and Grigoris J Karakoulas. 2003. Reinforcement learning for true adaptive traffic signal control. *Journal of Transportation Engineering* 129, 3 (2003), 278–285.
- [2] Itamar Arel, Cong Liu, T Urbanik, and AG Kohls. 2010. Reinforcement learning-based multi-agent system for network traffic signal control. *IET Intelligent Transport Systems* 4, 2 (2010), 128–135.
- [3] PG Balaji, X German, and Dipti Srinivasan. 2010. Urban traffic signal control using reinforcement learning agents. *IET Intelligent Transport Systems* 4, 3 (2010), 177–188.
- [4] Daniel S Bernstein, Robert Givan, Neil Immerman, and Shlomo Zilberstein. 2002. The complexity of decentralized control of Markov decision processes. *Mathematics of operations research* 27, 4 (2002), 819–840.
- [5] Yit Kwong Chin, Nurmin Bolong, Aroland Kiring, Soo Siang Yang, and Kenneth Tze Kin Teo. 2011. Q-learning based traffic optimization in management of signal timing plan. *International Journal of Simulation, Systems, Science & Technology* 12, 3 (2011), 29–35.
- [6] Seung-Bae Cools, Carlos Gershenson, and Bart D’Hooghe. 2013. Self-organizing traffic lights: A realistic simulation. In *Advances in applied self-organizing systems*. Springer, 45–55.
- [7] Samah El-Tantawy, Baher Abdulhai, and Hossam Abdelgawad. 2013. Multiagent reinforcement learning for integrated network of adaptive traffic signal controllers (MARLIN-ATSC): methodology and large-scale application on downtown Toronto. *IEEE Transactions on Intelligent Transportation Systems* 14, 3 (2013), 1140–1150.
- [8] K Nagel J Esser and M Rickert. 2000. *Large-scale traffic simulations for transportation planning*. World Scientific, Singapore.
- [9] Jakob N Foerster, Gregory Farquhar, Triantafyllos Afouras, Nantas Nardelli, and Shimon Whiteson. 2018. Counterfactual multi-agent policy gradients. In *Thirty-Second AAAI Conference on Artificial Intelligence*.
- [10] Wade Genders and Saeideh Razavi. 2016. Using a deep reinforcement learning agent for traffic signal control. *arXiv preprint arXiv:1611.01142* (2016).
- [11] Nikolas Geroliminis, Carlos F Daganzo, et al. 2007. Macroscopic modeling of traffic in cities. In *86th Annual Meeting of the Transportation Research Board, Washington, DC*.
- [12] Carlos Gershenson. 2004. Self-organizing traffic lights. *arXiv preprint nlin/0411066* (2004).
- [13] Federico Guerrini. 2014. Traffic Congestion Costs Americans \$124 Billion A Year, Report Says. *Forbes*, October 14 (2014).
- [14] Li Li, Yisheng Lv, and Fei-Yue Wang. 2016. Traffic signal timing via deep reinforcement learning. *IEEE/CAA Journal of Automatica Sinica* 3, 3 (2016), 247–254.
- [15] Xiaoyuan Liang, Xunsheng Du, Guiling Wang, and Zhu Han. 2018. Deep reinforcement learning for traffic light control in vehicular networks. *arXiv preprint arXiv:1803.11115* (2018).
- [16] Mengqi Liu, Jiachuan Deng, Ming Xu, Xianbo Zhang, and Wei Wang. 2017. Cooperative Deep Reinforcement Learning for Tra ic Signal Control. (2017).
- [17] Ying Liu, Lei Liu, and Wei-Peng Chen. 2017. Intelligent traffic light control using distributed multi-agent Q learning. *arXiv preprint arXiv:1711.10941* (2017).
- [18] Pablo Alvarez Lopez, Michael Behrisch, Laura Bieker-Walz, Jakob Erdmann, Yun-Pang Flötteröd, Robert Hilbrich, Leonhard Lücken, Johannes Rummel, Peter Wagner, and Evamarie Wießner. 2018. Microscopic Traffic Simulation using SUMO, In The 21st IEEE International Conference on Intelligent Transportation Systems. *IEEE Intelligent Transportation Systems Conference (ITSC)*. <https://elib.dlr.de/124092/>
- [19] Linsey McNew. 2014. Americans will waste \$2.8 trillion on traffic by 2030 if gridlock persists. (2014). <http://inrix.com/press-releases/americans-will-waste-2-8-trillion-on-traffic-by-2030-if-gridlock-persists/>
- [20] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. 2015. Human-level control through deep reinforcement learning. *Nature* 518, 7540 (2015), 529.
- [21] Seyed Sajad Mousavi, Michael Schukat, and Enda Howley. 2017. Traffic light control using deep policy-gradient and value-function-based reinforcement learning. *IET Intelligent Transport Systems* 11, 7 (2017), 417–423.
- [22] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. 1999. *The PageRank citation ranking: Bringing order to the web*. Technical Report. Stanford InfoLab.
- [23] Isaac Porche and Stéphane Lafontaine. 1999. Adaptive look-ahead optimization of traffic signals. *Journal of Intelligent Transportation System* 4, 3-4 (1999), 209–254.
- [24] Tabish Rashid, Mikayel Samvelyan, Christian Schroeder, Gregory Farquhar, Jakob Foerster, and Shimon Whiteson. 2018. QMIX: Monotonic Value Function Factorisation for Deep Multi-Agent Reinforcement Learning. In *Proceedings of the 35th International Conference on Machine Learning*, 4295–4304.
- [25] Peter Sunehag, Guy Lever, Audrunas Gruslys, Wojciech Marian Czarnecki, Vincius Zambaldi, Max Jaderberg, Marc Lanctot, Nicolas Sonnerat, Joel Z Leibo, Karl Tuyls, et al. 2017. Value-decomposition networks for cooperative multi-agent learning. *arXiv preprint arXiv:1706.05296* (2017).
- [26] Richard S Sutton and Andrew G Barto. 2018. *Reinforcement learning: An introduction*. MIT press.
- [27] Richard S Sutton, David A McAllester, Satinder P Singh, and Yishay Mansour. 2000. Policy gradient methods for reinforcement learning with function approximation. In *Advances in neural information processing systems*, 1057–1063.
- [28] Ardi Tampuu, Tambet Matiisen, Dorian Kodelja, Ilya Kuzovkin, Kristjan Korus, Juhan Aru, Jaan Aru, and Raul Vicente. 2017. Multiagent cooperation and competition with deep reinforcement learning. *PLoS one* 12, 4 (2017), e0172395.
- [29] Ming Tan. 1993. Multi-agent reinforcement learning: Independent vs. cooperative agents. In *Proceedings of the tenth international conference on machine learning*, 330–337.
- [30] Elise Van der Pol and Frans A Oliehoek. 2016. Coordinated deep reinforcement learners for traffic light control. *Proceedings of Learning, Inference and Control of Multi-Agent Systems (at NIPS 2016)* (2016).
- [31] Hua Wei, Guanjie Zheng, Huaxiu Yao, and Zhenhui Li. 2018. Intellilight: A reinforcement learning approach for intelligent traffic light control. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 2496–2505.
- [32] Cathy Wu, Aboudy Kreidieh, Kanaad Parvate, Eugene Vinitsky, and Alexandre M Bayen. 2017. Flow: Architecture and benchmarking for reinforcement learning in traffic control. *arXiv preprint arXiv:1710.05465* (2017).
- [33] Jiachen Yang, Alireza Nakhaei, David Isele, Hongyuan Zha, and Kikuo Fujimura. 2018. CM3: Cooperative Multi-goal Multi-stage Multi-agent Reinforcement Learning. *arXiv preprint arXiv:1809.05188* (2018).



Dynamic patterns and predictors of hydroxychloroquine nonadherence among Medicaid beneficiaries with systemic lupus erythematosus



Candace H. Feldman, MD, ScD^{a,b,*}, Jamie Collins, PhD^c, Zhi Zhang, MS^a, S.V. Subramanian, PhD^b, Daniel H. Solomon, MD, MPH^a, Ichiro Kawachi, MB.ChB, PhD^{b,1}, Karen H. Costenbader, MD, MPH^{a,1}

^a Division of Rheumatology, Immunology and Allergy, Department of Medicine, Brigham and Women's Hospital, Boston, MA

^b Department of Social and Behavioral Sciences, Harvard T. H. Chan School of Public Health, Boston, MA

^c Department of Orthopedic Surgery, The Orthopaedic and Arthritis Center for Outcomes Research (OrACORE), Brigham and Women's Hospital, Boston, MA

ARTICLE INFO

Keywords:
Adherence
Hydroxychloroquine
Systemic lupus erythematosus
Pharmacoepidemiology
Observational study
Health services research

ABSTRACT

Background: Hydroxychloroquine (HCQ) is the standard of care medication for most SLE patients, however nonadherence is common. We investigated longitudinal patterns and predictors of non-adherence to HCQ in a U.S. SLE cohort of HCQ initiators.

Methods: We used Medicaid data from 28 states to identify adults 18–65 years with prevalent SLE. We included HCQ initiators following ≥6 months without use, and required ≥1 year of follow-up after first dispensing (index date). We used the proportion of days covered (PDC) to describe overall HCQ adherence (< 80% = nonadherent) and novel group-based trajectory models (GBTM) to examine monthly patterns (< 80% of days/month covered = nonadherent), during the first year of use. Multivariable multinomial logistic regression models were used to examine predictors of nonadherence.

Results: We identified 10,406 HCQ initiators with SLE. Mean age was 38 (± 12) years, 94% were female, 42% black, 31% white; 85% had a mean PDC < 80%. In our 4-group GBTM, 17% were persistent adherers, 36% persistent nonadherers, and 47% formed two dynamic patterns of partial adherence. Adherence declined for most patients over the first year. Compared to persistent adherers, the odds of nonadherence were increased for blacks and Hispanics vs. whites and for younger ages vs. older; increased SLE-related comorbidities were associated with reduced odds of nonadherence for persistent nonadherers (0.95, 95% CI: 0.91–0.99).

Conclusions: Among HCQ initiators with SLE, we observed poor adherence which declined for most over the first year of use. HCQ adherence is a dynamic behavior and further studies of associated predictors, outcomes, and interventions should reflect this.

© 2018 Elsevier Inc. All rights reserved.

Introduction

Medication nonadherence is a serious problem among patients with systemic lupus erythematosus (SLE); less than half of patients

adhere to their SLE-related medications as prescribed [1]. Clinical and epidemiologic factors unique to SLE may increase nonadherence including cognitive and psychological manifestations, high disease burden among lower socioeconomic status groups, the complexity and toxicity of the medication regimens, and SLE disease activity fluctuations. Hydroxychloroquine (HCQ) is considered the backbone of SLE therapy regardless of disease severity, and it is now standard of care for all SLE patients to take HCQ continuously beginning at the time of diagnosis [2–4]. HCQ use is disease stabilizing and associated with fewer disease flares, reduced disease activity overall and less organ damage accrual [2,4–7]. Medically indicated discontinuation is uncommon with the exception of evidence of retinal toxicity, which results in most cases from cumulative exposure and occurs in 4–7.5% of patients taking HCQ for 10 years and in < 1% during the first 5–7 years [7,8].

Abbreviations: SLE, systemic lupus erythematosus; HCQ, hydroxychloroquine; PDC, proportion of days covered; GBTM, group-based trajectory models; MAX, Medicaid Analytic eXtract; ED, emergency department.

Funding: This study was funded by the Rheumatology Research Foundation Investigator Award (CH Feldman), NIH 1K23 AR071500-01 (CH Feldman), NIH NIAMS R01 057327 (KH Costenbader), K24 AR066109 (KH Costenbader), and K24 AR055989 (DH Solomon).

* Corresponding author at: Division of Immunology, Allergy and Immunology, Brigham and Women's Hospital, 60 Fenwood Road, Office 6016P, Boston, MA 02115.

E-mail address: cfeldman@bwh.harvard.edu (C.H. Feldman).

¹These authors contributed equally.

To date, most studies of HCQ adherence are small, cross-sectional, and based in academic cohorts. Moreover, they rely on one-time often self-reported measures of adherence failing to capture the dynamic nature of adherence behavior over time. Conflicting results regarding risk factors for nonadherence, and physicians' inability to accurately predict who is likely to non-adhere, make it difficult to know who to target and how to intervene [9,10]. In addition, most studies included prevalent users of HCQ and therefore conflate potentially different risk factors for nonadherence among patients initiating HCQ with those who have been taking it for years. We therefore aimed to use nationwide data on patients enrolled in Medicaid, the federal-state public health insurance for low-income individuals, to describe longitudinal patterns and predictors of HCQ adherence among SLE patients newly receiving this medication. To define distinct, dynamic HCQ adherence patterns, we used a well-described but novel method, group-based trajectory models (GBTM). GBTMs have been used in psychology and other social sciences, to model underlying longitudinal patterns where there are repeated measures available for individuals that may change over time [11,12]. In the chronic disease literature, there are a few studies that use GBTM to describe patterns of adherence behavior and the method has been shown to better capture changes in adherence over time than standard composite measures such as the proportion of days covered (PDC) [13–16]. To our knowledge, GBTMs have not been previously used to describe adherence among patients with SLE. We hypothesized that GBTMs would demonstrate distinct patterns of declining adherence over the first year of use and certain sociodemographic (e.g., young age and black race) and disease-related (e.g., absence of lupus nephritis) characteristics would predict patterns of sustained nonadherence.

Material and methods

Patient cohort

We used the Medicaid Analytic eXtract (MAX) from the 29 most populated U.S. states from 2000–2010. HCQ dispensing data were unavailable in MAX for Medicaid beneficiaries living in Ohio and therefore this state was excluded, leaving 28 states in our analysis. MAX includes all billing claims, healthcare utilization, drug-dispensing data, and demographic information for Medicaid beneficiaries. We identified patients aged 18–65 years with prevalent SLE based on ≥2 International Classification of Diseases, Ninth Revision (ICD-9) codes for SLE (710.0) from hospital discharge diagnoses or physician visit claims ≥30 days apart, and HCQ dispensing within 365 days of a SLE ICD-9 code. In prior studies using MAX, we required ≥3 ICD-9 codes for SLE; however, in this study we aimed to increase our ability to capture all patients with new onset SLE who initiated HCQ and therefore employed ≥2 codes for SLE plus one code for HCQ to accomplish this [17]. This algorithm with ≥2 ICD-9 codes and a related medication has been validated both in the rheumatoid arthritis literature, as well as in electronic health record-based analyses among SLE patients with PPVs ranging from 77% to 89% [18,19]. In addition, our interest here was HCQ adherence patterns and not SLE-associated outcomes, as it had been for the prior studies [20,21]. We restricted our cohort to SLE patients with ≥183 days of continuous enrollment prior to first HCQ dispensing (index date) with no use of HCQ during this 183-day period. We restricted our cohort to incident users of HCQ to avoid potential bias from depletion of susceptibles [22]. Prevalent users of HCQ may be different from incident users because they remained on the drug (e.g., they did not have side effects or other reasons to discontinue or become nonadherent early on in use) and therefore we did not want to combine these

groups in our analyses. We included patients who had ≥365 days of continuous follow-up after the index date. We excluded those with no dispensing data ($N = 233$), and those who were hospitalized for the entire duration of follow-up ($N = 18$). Additionally, we excluded patients who were missing zip code data as median household income was considered a potentially important covariate ($N = 253$).

Adherence measures

We assessed adherence in two ways using prescription refill data. Refill adherence has been shown to be a valid source of adherence information in population-based studies when directly observed methods are not possible [23,24]. Prior studies demonstrate significant associations between refill adherence and other measures of adherence, as well as with serum drug levels and physiologic drug effects [23,25–28]. Using prescription refill data, we calculated the overall proportion of days covered (PDC) for the 365-day follow-up period beginning at the index date. We calculated the PDC as the number of days covered divided by 365 days, multiplied by 100. We subtracted hospitalized days from the numerator and the denominator. In keeping with the chronic disease medication adherence literature [29], PDC ≥ 80% was considered adherent. In addition, we created a 12-month diary for each patient assigning a binary variable (0 (nonadherent) or 1 (adherent)) for each 30-day period indicating whether that period had ≥ 80% coverage (24 of the 30 days) with HCQ. We chose to use 30 days because 91% of our cohort received 30-day supplies or less of HCQ.

Covariates

We assessed covariates during the 183-day baseline period prior to and including the index date, unless otherwise specified. Demographic factors included age at the index date, sex, race/ethnicity (White, Black or African American, Hispanic or Latino, Asian (including Native Hawaiian and Pacific Islander), American Indian/Alaska Native, and other), region and state of residence at the index date. We included zip code median household income as a proxy for individual socioeconomic status using American Community Survey data (2006–2010) [30]. We assessed comorbidities including thromboembolism, pulmonary disease, chronic kidney disease, liver disease, cardiovascular disease, cerebrovascular disease, alcoholism, substance abuse, obesity or cancer/hematologic malignancy using ≥ 1 ICD-9 codes. For diabetes, we required an ICD-9 code for diabetes or the prescription of a diabetes-related medication. To determine smoking status, we used ≥ 1 ICD-9 code, CPT code for smoking cessation counseling, or dispensing of smoking cessation-related medications. We used the SLE risk adjustment index, which has been shown to be a better predictor of inpatient mortality among SLE patients than the Charlson comorbidity index, as a measure of SLE-related comorbidities [31]. We determined lupus nephritis using ≥ 2 ICD-9 hospital discharge diagnosis codes or physician billing claims for nephritis, proteinuria and/or renal failure occurring on or after one SLE diagnosis code [32]. We included antidepressant use as a marker of depression given the low positive predictive value of depression-related claims [33].

As a marker of disease activity, we included number of laboratory tests for anti-dsDNA, BUN, creatinine, urinalysis and sediment, complement, ESR, and CRP. We included measures of healthcare utilization (number of emergency department (ED) visits, outpatient visits and hospitalizations, and number of hospitalized days), as well as preventive care using CPT codes for influenza and pneumococcal vaccinations. We assessed the number of other medications filled on the index date of new HCQ use, and HCQ dispensing quantity. In the 183 days prior to and including the index date, we obtained number of drugs,

ever/never use of immunosuppressive medications (methotrexate, sulfasalazine, cyclophosphamide, azathioprine, mycophenolate mofetil, rituximab, tacrolimus, or leflunomide), as well as corticosteroid use (mean daily prednisone-equivalent dose), anticoagulant use, nonsteroidal anti-inflammatory medications and selective and non-selective cyclooxygenase-2 inhibitors.

Statistical methods

We calculated the overall PDC during the 365-day period following the index date. We used our binary indicators of adherence (0 (< 80%), 1 ($\geq 80\%$)) for each 30-day period for the 365 days of follow-up to develop group-based trajectory models (GBTM) to classify patients by their HCQ adherence. GBTM has been shown to be the optimal technique for identifying latent patterns behind longitudinal data [13,34]. First, a multinomial logistic regression model is used with no predictors and an intercept for each group to model the probability of each SLE patient's probability of belonging to each adherence trajectory based on his/her own adherence pattern [13]. The GBTM accounts for repeated monthly measures of adherence for each individual and treats them as independent conditional on the trajectory group [13]. Each individual is then assigned to the group with the highest probability of membership. Similar to a prior study, we modeled time in various forms (linear, quadratic, and cubic) and found that the cubic terms best fit the adherence patterns uncovered in our data [13]. We estimated trajectory models ranging from 2 to 6 groups. We did this because we aimed to find a GBTM that had good model fit but was not too complex and allowed for a more nuanced understanding of nonadherence patterns beyond a dichotomized comparison of overall "adherers" and "nonadherers." We examined mean posterior probability values, or the average conditional probability of having been assigned to each trajectory group given a person's adherence, with a probability of more than 80 percent for each group within each model considered to be acceptable. We compared model fit using Bayesian information criterion (BIC) with lower values considered preferable [35]. We looked for a reasonable distribution of subjects across the groups, and one clearly adherent trajectory (mean PDC $\geq 80\%$) for comparison. In choosing the most appropriate model, we aimed to balance what GBTM expert D.S. Nagin described as the "sometimes competing objectives of model parsimony and capturing the distinctive features of the data." [35].

We then determined the association between baseline socio-demographic, clinical and healthcare utilization-related characteristics and the probability of trajectory group membership. We used multinomial logistic regression models to examine the odds of belonging to a nonadherent trajectory compared to the persistently adherent trajectory for the predictors including covariates during the 183 days prior to the index date, calendar year, and state of residence at the index date.

We conducted additional analyses comparing the two trajectories that started off similar and then diverged 4–5 months after the index date. We used month 5 as our new index date and updated baseline covariates to include the period between the original index date and the fifth month of follow-up. We used multi-variable logistic regression models to assess predictors of the two trajectories. We also compared monthly utilization between months 4 and 7, the time-period surrounding the point where the two curves diverged.

All analyses were conducted using SAS 9.4 (Cary, NC) and we used "Proc Traj," an add-on package to base SAS to create our GBTM. Data were obtained from the Centers for Medicare and Medicaid Services (CMS) through a Data Use Agreement and in accordance with CMS policies, all cell sizes < 11 are suppressed. The Partners Healthcare Institutional Review Board approved this study.

Results

We identified 10,406 individuals with SLE who were new users of HCQ, had complete HCQ dispensing data, and 365 days of follow-up beginning at the date of HCQ initiation. The mean \pm SD age was 37.7 ± 11.8 years, 94% were female, 42% were black, 31% white, 20% Hispanic (Table 1). During the baseline period, 10% had ICD-9 codes consistent with lupus nephritis, 27% with cardiovascular disease, 29% received an antidepressant medication, and 59% received corticosteroids. During the 365-day follow-up period, the overall mean \pm SD PDC was $42\% \pm 29$; 15% of patients ($N = 1575$) had a composite PDC $\geq 80\%$.

Group-based trajectory model (GBTM)

We found that a four-group trajectory model allowed for the most nuanced exploration of nonadherence while also providing a good fit for our data (Figure). The five- and six-group models had larger BICs, group posterior probabilities less than 80 percent, and unbalanced sample size distributions. The BICs for the two- and three-group models were smaller. However, the two-group model did not have an adherent trajectory with a mean PDC $\geq 80\%$. To both uphold model parsimony and understand unique features of nonadherent patterns in depth, we chose the four-group model over the three-group model. In keeping with Nagin's principles, the BICs were similar, sample sizes were balanced, and both had all group posterior probabilities greater than 80 percent, however the four-group model uncovered distinctive patterns that enabled further exploration [35] (Supplemental Figure 1).

The four-group model revealed four distinct patterns (Figure). Group 1, which includes 36% of the cohort, were persistent non-adherers with few if any refills of HCQ after the initial dispensing. Group 4, which included 17% of the cohort, were persistent adherers, with, on average, $\geq 80\%$ of days covered for nearly all months over the course of the year of follow-up with a slightly declining trend from months 10 to 12. Groups 2 and 3 were intermediate nonadherers with more dynamic patterns than groups 1 and 4. The trajectories for Groups 2 and 3 were similar until month 5 when they diverged; Group 3 improved slightly and then reached a plateau whereas Group 2 became nearly completely nonadherent for the remainder of follow-up. Except for Group 3, adherence declined by the end of the year of follow-up compared with the first 90 days. For Group 1, the overall mean \pm SD PDC was $15\% \pm 10$, for Group 2, $37\% \pm 15$, for Group 3, $57\% \pm 15$, and for Group 4, $88\% \pm 8$. There were no individuals in Group 1 with mean PDCs $\geq 80\%$, and 1439 (84%) in Group 4.

Baseline characteristics for the four trajectories are presented in Table 2. The mean age was highest among the persistent adherers (Group 4) and lowest among persistent nonadherers (Group 1) ($p < 0.001$). The highest percentage of individuals in Groups 1, 2, and 3 were black compared to Group 4 where the highest percentage were white. The median household income was similar across groups with slightly higher income among individuals in Group 4 compared to Group 1 ($p = 0.01$). The SLE risk adjustment index was highest in Group 4 suggesting a higher burden of SLE-related comorbidities ($p < 0.001$). Similarly, the mean number of medications dispensed ($p < 0.001$) and the mean daily prednisone-equivalent dose were both higher in Group 4 compared to Group 1 suggesting a more ill population ($p < 0.001$).

Trajectory predictors

We estimated multinomial logistic regression models to examine predictors of the different trajectories with Group 4 (persistent adherers) as the reference. We found increased odds of belonging to all three nonadherent trajectories for individuals aged 18–34

Table 1

Baseline characteristics of new users of HCQ with SLE enrolled in Medicaid

Baseline characteristics ^a	HCQ new users (N = 10,406)
Age—mean ± SD	37.7 ± 11.8
Age group—N (%)	
18–34 years	4614 (44)
35–50 years	3951 (38)
51–65 years	1841 (18)
Female—N (%)	9800 (94)
Race/ethnicity	
Black	4365 (42)
White	3239 (31)
Hispanic	2047 (20)
Asian	400 (4)
American Indian/Alaska Native	121 (1)
Other	234 (2)
Region	
Northeast	2507 (24)
Midwest	1607 (15)
South	3789 (36)
West	2503 (24)
Median household income ^b —mean ± SD	4.5 ± 1.7
SLE risk adjustment index—mean ± SD	1.0 ± 1.9
Comorbidities—N (%)	
Substance abuse	154 (1)
Alcoholism	57 (1)
Malignancy	235 (2)
Cardiovascular disease	2818 (27)
Cerebrovascular disease	311 (3)
Chronic kidney disease	73 (1)
Diabetes mellitus	972 (9)
Chronic liver disease	342 (3)
Chronic lung disease	1167 (11)
Lupus nephritis	1059 (10)
Obesity	240 (2)
Thromboembolic disease	359 (3)
Smoking	627 (6)
Antidepressant use	3019 (29)
Preventive care—N (%)	
Influenza vaccine	184 (2)
Pneumococcal vaccine	63 (1)
Immunosuppressive medication use—N (%)	
Azathioprine	540 (5)
Cyclophosphamide	38 (0.4)
Leflunomide	66 (1)
Methotrexate	601 (6)
Mycophenolate mofetil	378 (4)
Sulfasalazine	88 (1)
Tacrolimus	55 (1)
Corticosteroids	
Ever use—N (%)	6160 (59)
Mean daily prednisone-equivalent dose ± SD	2.9 mg ± 17 Median: 0 (0, 2.8)
Mean number of medications—mean ± SD	4.1 ± 3.4
Healthcare utilization	
ED visits—median (25, 75)	0 (0, 1)
Hospitalizations—median (25, 75)	0 (0, 1)
Outpatient visits—median (25, 75)	2 (0, 6)
Hospitalized days—mean ± SD	3.9 ± 11

^a Determined from the 183 days prior to and including the index date (the date of first HCQ dispensing).^b Determined at the zip code level; mean ± SD divided by 10,000.

years and 35–50 years, compared to those 51–65 years (Table 3). We similarly found increased odds of belonging to Groups 1, 2, or 3 compared to Group 4 associated with black race and Hispanic ethnicity, compared to white. We found reduced odds (OR = 0.64, 95% CI: 0.47–0.87) of belonging to Group 1 vs. Group 4 (persistent adherers) among Asians compared to whites.

In terms of comorbidities, we found increased odds of belonging to Group 1 (persistent nonadherers) vs. 4 (persistent adherers) associated with diabetes (OR = 1.25, 95% CI: 1.01–1.56) and decreased odds for each unit increase in the SLE risk adjustment index (0.95, 95% CI: 0.91–0.99). There were reduced odds of belonging to Group 3 vs. 4 associated with lupus nephritis (OR = 0.71, 95% CI: 0.53–0.96). The odds of belonging to nonadherent Groups 1, 2, and 3 vs. Group 4 were inversely related to the total number of medications filled. There were increased odds of belonging to nonadherent Groups 1 and 2 vs. 4 associated with antidepressant use.

We used multivariable logistic regression to examine predictors of belonging to Group 2 vs. Group 3 beginning at month 5, the point at which the curves diverged and Group 2 became nearly completely nonadherent while Group 3 plateaued (Table 4). We found increased odds of belonging to Group 2 associated with younger age (OR = 1.20, 95% CI: 1.01–1.44) and the use of antidepressants (OR = 1.25, 95% CI: 1.10–1.43). We examined healthcare utilization separately by month for months 4 through 7 (Table 5) and found that beginning at month 4, patients in Group 3 had more hospitalizations, and beginning at month 5, longer hospitalizations compared to those in Group 2. We observed a trend toward more outpatient visits for Group 3 compared to Group 2 during months 6 and 7.

Discussion

In this longitudinal study of Medicaid beneficiaries with SLE, adherence among HCQ initiators was poor starting one month after the first dispensing, and for most patients, adherence declined over the first year of use. While prior studies demonstrated that nonadherence among SLE patients is common, nonadherence among SLE patients enrolled in Medicaid is even more pronounced [10,36,37]. Our model revealed a group of persistent nonadherers, which comprised 36% of our cohort, who had very few HCQ refills after the initial dispensing. We identified a small group of persistent adherers (Group 4, 17%), although even this group also experienced a decline in adherence beginning at 9 months.

In contrast to prior studies, which either measured adherence cross-sectionally or using a composite measure, we could explore

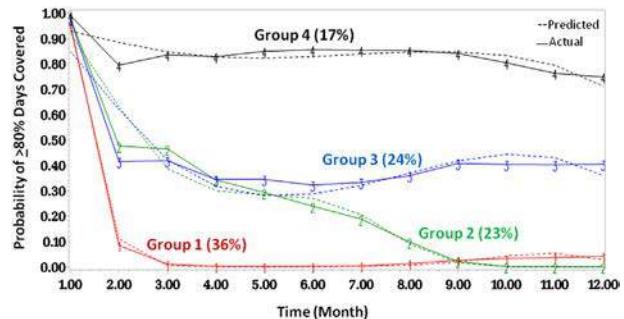


Fig. 4. Four-group trajectory model of adherence patterns for new HCQ users with SLE; Group 1 are persistent nonadherers, Group 4 are persistent adherers.

Table 2

Baseline characteristics by trajectory group (N = 10,406)

Baseline characteristics ^a	Group 1 (persistent nonadherers)	Group 2 (intermediate nonadherers)	Group 3 (intermediate nonadherers)	Group 4 (persistent adherers)
N (%)	3772 (36)	2431 (23)	2481 (24)	1722 (17)
Age—mean ± SD	36.7 ± 11.6	37.1 ± 11.7	37.8 ± 11.8	40.4 ± 12.2
Age group—N (%)				
18–34 years	1808 (48)	1120 (46)	1067 (43)	619 (36)
35–50 years	1388 (37)	923 (38)	975 (39)	665 (39)
51–65 years	576 (15)	388 (16)	439 (18)	438 (25)
Female—N (%)	3568 (95)	2301 (95)	2333 (94)	1598 (93)
Race/ethnicity				
Black	1694 (45)	1069 (44)	1062 (43)	540 (31)
White	1166 (31)	675 (28)	717 (29)	681 (40)
Hispanic	692 (18)	518 (21)	513 (21)	324 (19)
Asian	99 (3)	90 (4)	100 (4)	111 (6)
AI/AN	46 (1)	27 (1)	28 (1)	20 (1)
Other	75 (2)	52 (2)	61 (2)	46 (3)
Region				
Northeast	843 (22)	579 (24)	626 (25)	459 (27)
Midwest	627 (17)	357 (15)	357 (14)	266 (15)
South	1455 (39)	922 (38)	867 (35)	545 (32)
West	847 (22)	573 (24)	631 (25)	452 (26)
Median household income ^b —mean ± SD	4.4 ± 1.7	4.5 ± 1.6	4.5 ± 1.7	4.6 ± 1.7
SLE risk adjustment index—mean ± SD	0.9 ± 1.8	1.0 ± 1.8	1.1 ± 2.0	1.3 ± 2.2
Comorbidities—N (%)				
Substance abuse	71 (2)	33 (1)	27 (1)	23 (1)
Alcoholism	23 (1)	11 (0.5)	17 (1)	NR
Malignancy	78 (2)	48 (2)	63 (3)	46 (3)
Cardiovascular disease	942 (25)	658 (27)	709 (29)	509 (30)
Cerebrovascular disease	99 (3)	72 (3)	84 (3)	56 (3)
Chronic kidney disease	27 (1)	15 (1)	20 (8)	11 (1)
Diabetes mellitus	314 (8)	221 (9)	242 (10)	195 (11)
Chronic liver disease	112 (3)	81 (3)	85 (3)	64 (4)
Chronic lung disease	402 (11)	299 (12)	269 (11)	197 (11)
Lupus nephritis	366 (10)	244 (10)	252 (10)	197 (11)
Obesity	90 (2)	45 (2)	62 (2)	43 (2)
Thromboembolic disease	109 (3)	84 (3)	84 (3)	82 (5)
Smoking	215 (6)	149 (6)	148 (6)	115 (7)
Antidepressant use	1045 (28)	739 (30)	685 (28)	550 (32)
Preventive care—N (%)				
Influenza vaccine	66 (2)	32 (1)	48 (2)	38 (2)
Pneumococcal vaccine	21 (1)	11 (0.5)	17 (1)	14 (1)
Immunosuppressive medication use—N (%)				
Azathioprine	160 (4)	107 (4)	151 (6)	122 (7)
Cyclophosphamide	11 (0.3)	NR	NR	NR
Leflunomide	20 (0.5)	15 (1)	11 (0.4)	20 (1)
Methotrexate	204 (5)	138 (6)	156 (6)	103 (6)
Mycophenolate mofetil	122 (3)	87 (4)	99 (4)	70 (4)
Sulfasalazine	34 (1)	19 (8)	20 (8)	15 (1)
Tacrolimus	18 (0.5)	13 (1)	NR	15 (1)
Corticosteroids				
Ever use—N (%)	2108 (56)	1475 (61)	1521 (61)	1056 (61)
Mean daily prednisone-equivalent dose ± SD	2.5 mg ± 5.8 Median 0 (0, 2.4)	2.9 mg ± 6.5 Median 0 (0, 2.9)	2.8 mg ± 6.0 Median 0 (0, 3)	4 mg ± 40 Median 0 (0, 3.3)
HCQ Prescription \leq 30 days—N (%)	3557 (94)	2131 (88)	2270 (91)	1490 (87)
Mean number of medications—mean ± SD	3.7 ± 3.2	4.2 ± 3.4	4.2 ± 3.3	5.3 ± 3.9
Healthcare utilization				
ED visits—median (25, 75)	0 (0, 1) Mean 0.95 ± 2.4	0 (0, 1) Mean 0.91 ± 2.0	0 (0, 1) Mean 0.82 ± 1.9	0 (0, 1) Mean 0.76 ± 1.9

Table 2 (continued)

Baseline characteristics ^a	Group 1 (persistent nonadherers)	Group 2 (intermediate nonadherers)	Group 3 (intermediate nonadherers)	Group 4 (persistent adherers)
Inpatient—median (25, 75)	0 (0, 1) Mean 0.57 ± 1.1	0 (0, 1) Mean 0.60 ± 1.2	0 (0, 1) Mean 0.60 ± 1.1	0 (0, 1) Mean 0.61 ± 1.2
Outpatient—median (25, 75)	2 (0, 6) Mean 3.9 ± 4.8	2 (0, 6) Mean 3.9 ± 4.9	2 (0, 6) Mean 3.7 ± 4.8	2 (0, 7) Mean 4.0 ± 5.0
Hospitalized days—mean ± SD	3.5 ± 9.5	4.2 ± 12.7	3.9 ± 10.0	4.3 ± 11.0

^a Determined from the 183 days prior to and including the index date (the date of first HCQ dispensing).

^b Determined at the zip code level; mean ± SD divided by 10,000.

the nuances of adherence patterns over the first year of use, which was especially relevant for the intermediate nonadherers (Groups 2 and 3). We found that using the PDC, a commonly used composite measure, we would have misclassified 136 patients with nonadherent patterns as adherent and 283 patients with persistently adherent patterns as nonadherent. We found that the 5-month mark represented a critical juncture that may represent a clinical opportunity to intervene before adherence worsens among these “undecided” groups. We found that patients who plateaued had more frequent and longer hospitalizations suggesting both that they were more seriously ill, and

that they likely had more interactions with the healthcare system to have their medications renewed. We also observed a trend toward more outpatient visits in this group suggesting that sustained access to outpatient care may increase the likelihood a patient continues the medication he/she is prescribed. Five months might also be the point at which patients feel that they have adequately trialed the medication and if there is no symptomatic improvement, they discontinue. With the growing body of literature suggesting long-term preventive effects from HCQ, increased provider and patient education at this juncture may be beneficial [2–4,6,7,38].

Table 3

Multinomial logistic regression model of the odds of being in the Group 1 (persistent nonadherers), Group 2 or 3 (intermediate nonadherers) trajectories compared with being in the Group 4 trajectory (persistent adherers, reference)

Baseline characteristics N = 10,406	Group 1 (persistent nonadherers) OR (95% CI)	Group 2 (intermediate nonadherers) OR (95% CI)	Group 3 (intermediate nonadherers) OR (95% CI)
Age group			
18–34 years	1.66 (1.39–1.98)	1.66 (1.38–2.02)	1.42 (1.17–1.71)
35–50 years	1.33 (1.13–1.57)	1.38 (1.16–1.66)	1.32 (1.11–1.58)
51–65 years	Ref.	Ref.	Ref.
Male (female = ref)	0.87 (0.68–1.10)	0.85 (0.66–1.11)	0.88 (0.68–1.14)
Race/ethnicity			
White	Ref.	Ref.	Ref.
Black	1.74 (1.49–2.04)	1.95 (1.65–2.31)	1.81 (1.53–2.14)
Hispanic	1.40 (1.16–1.68)	1.66 (1.36–2.02)	1.51 (1.24–1.83)
Asian	0.64 (0.47–0.87)	0.92 (0.67–1.26)	0.86 (0.63–1.17)
AI/AN	1.05 (0.60–1.84)	1.15 (0.62–2.12)	1.11 (0.61–2.03)
Median household income	0.99 (0.96–1.03)	1.01 (0.96–1.05)	1.00 (0.96–1.04)
SLE risk adjustment index	0.95 (0.91–0.99)	0.96 (0.92–1.00)	0.99 (0.95–1.04)
Diabetes mellitus	1.25 (1.01–1.56)	1.17 (0.93–1.49)	1.18 (0.94–1.48)
Lupus nephritis	1.06 (0.80–1.41)	0.90 (0.66–1.22)	0.71 (0.53–0.96)
Antidepressant use (Never = ref)	1.18 (1.02–1.37)	1.30 (1.12–1.52)	1.11 (0.95–1.29)
Corticosteroids use (Never = ref)	0.87 (0.76–1.00)	0.99 (0.86–1.14)	1.07 (0.93–1.23)
Number of medications	0.90 (0.88–0.92)	0.94 (0.92–0.96)	0.92 (0.90–0.94)
Healthcare utilization			
ED visits	1.05 (1.01–1.08)	1.03 (0.99–1.07)	1.02 (0.98–1.06)
Hospitalizations	1.12 (1.03–1.22)	1.05 (0.96–1.14)	1.06 (0.98–1.16)
Outpatient visits	1.00 (0.98–1.01)	0.99 (0.97–1.01)	1.00 (0.98–1.01)

Model additionally adjusted for U.S. state, geographic region, calendar year at index date, index date HCQ dispensing amount, laboratory tests, pain medications, preventive care (influenza vaccine and pneumococcal vaccine), immunosuppressive medication use, comorbidities (substance abuse, alcoholism, malignancy, cardiovascular disease, cerebrovascular disease, chronic kidney disease, obesity, thromboembolic disease, chronic lung disease, and smoking), and mean daily corticosteroid dose. All variables were determined during the 183 days prior to and including the index date. Group 4 (persistent adherers) is their reference. All bold values are statistically significant. OR = odds ratio, CI = confidence interval.

Table 4

Multivariable logistic regression model comparing Group 2 ($N = 2431$, declining adherence) to Group 3 ($N = 2481$, plateaued adherence, reference) at the point of divergence^a

Predictors ^a	Group 2 ^b	
	Odds ratio	(95% CI)
Age group		
18–34 years	1.20 (1.01–1.44)	
35–50 years	1.07 (0.90–1.27)	
51–65 years	Ref	
Male (female = ref)	0.97 (0.75–1.25)	
Race/ethnicity		
White	Ref	
Black	1.09 (0.93–1.27)	
Hispanic	1.13 (0.94–1.36)	
Asian	1.09 (0.79–1.51)	
AI/AN	1.07 (0.61–1.88)	
Median household income	1.01 (0.97–1.05)	
SLE risk adjustment index	0.96 (0.92–1.01)	
Diabetes mellitus	1.01 (0.81–1.26)	
Lupus nephritis	1.32 (0.99–1.74)	
Antidepressant use (never = ref)	1.16 (1.01–1.34)	
Corticosteroids use (never = ref)	0.93 (0.82–1.06)	
Number of medications	1.02 (1.00–1.04)	
Healthcare utilization		
ED visits	1.01 (0.99–1.05)	
Hospitalizations	0.99 (0.92–1.07)	
Outpatient visits	0.99 (0.98–1.01)	
Number of laboratory tests		
ESR	1.09 (0.79–1.23)	
BUN	1.24 (1.04–1.49)	
Creatinine	0.96 (0.84–1.09)	
Complement (C3 or C4)	0.94 (0.82–1.06)	

Model is additionally adjusted for U.S. state, geographic region, calendar year at index date, index date HCQ dispensing amount, additional laboratory tests, pain medications, preventive care (influenza vaccine and pneumococcal vaccine), immunosuppressive medication use, comorbidities (substance abuse, alcoholism, malignancy, cardiovascular disease, cerebrovascular disease, chronic kidney disease, obesity, thromboembolic disease, chronic lung disease, and smoking), and mean daily corticosteroid dose. All bold values are statistically significant.

^a Predictors from 6 months prior to first HCQ dispensing and updated through month 4 of follow-up; nonadherence patterns assessed from months 5 through 12.

^b Group 3 is the reference.

Interestingly, we did not find many strong predictors within our available set of covariates that were significantly associated with declining adherence (Group 2) versus plateaued adherence (Group 3) at 5 months. We did not find an association with demographic factors which suggests that while age and race/

ethnicity might contribute initially to who is likely to be a persistent adherer or nonadherent, these factors may not play an important role in determining who continues to be modestly adherent versus who discontinues after an initial period. We did find an association with increased antidepressant use among those with declining adherence. Prior studies similarly show depression as a risk factor for poorer adherence and it is possible that among patients with depression the threshold to discontinue a medication that they may not see a tangible benefit after a period of time may be lower [39].

In order to investigate whether patterns of adherence are distinct by drug, we have separately used group-based trajectory models to examine adherence to azathioprine and mycophenolate mofetil also within the Medicaid SLE population [40]. Interestingly, while we similarly noted high rates of nonadherence to both drugs with fewer than 20% of individuals characterized as adherent to either drug, we found that each drug had a pattern of adherence that differed from that of HCQ. While both drugs had subsets of the population who were persistent adherers and persistent non-adherers, the paths of the intermediate nonadherers were distinct. This suggests that while there may be certain patients who will adhere or will not adhere consistently, for those in a more “undecided” category, characteristics of the drug itself, such as side effect profile, may play a role in adherence behavior. In addition, the patients receiving azathioprine and mycophenolate mofetil are likely sicker than those receiving HCQ, and this, as well as ongoing disease activity, may also contribute to adherence behavior. Overall, our finding of high rates of nonadherence among SLE patients is in line with prior studies that repeatedly show that more than half of SLE patients are nonadherent regardless of the adherence assessment method or the population under investigation [41]. Our findings are also consistent with adherence estimates for other chronic disease medications both among Medicaid beneficiaries and among the commercially insured [42,43].

Previous studies have highlighted several potential predictors for nonadherence including black race, increased comorbidities, depression, and polypharmacy [39,44,45]. We found that younger age, black race, Hispanic ethnicity, and antidepressant medication use increased the odds of nonadherence. However, in our cohort, corticosteroid use, polypharmacy, and higher SLE risk adjustment scores, all associated with increased SLE severity, reduced the odds of persistent nonadherence. Our findings suggest that patients who may have more active and severe SLE may be more likely to adhere to their HCQ. It is possible that among patients with milder SLE, while it is the standard of care to continue HCQ in all SLE patients unless contraindicated, the patients or physicians may have felt the medication was unnecessary. Overall, it is important to note that the magnitude of all of the statistically significant odds ratios was small. This reaffirms findings from prior studies that note that adherence behavior is complex and not readily predictable by a defined set of measurable variables.

There were limitations to this work. First, we used dispensing data to infer adherence however, filling a medication does not guarantee that a medication was taken. While 91% of our cohort

Table 5

Healthcare utilization for Group 2 ($N = 2431$, declining adherence) and Group 3 ($N = 2481$, plateaued adherence) in months 4–7

Healthcare utilization	Month 4		Month 5		Month 6		Month 7	
	Group 2	Group 3						
Outpatient visits—mean \pm SD	8.4 \pm 7.0	8.4 \pm 6.8	9.3 \pm 7.7	9.3 \pm 7.5	10.1 \pm 8.2	10.2 \pm 8.1	10.9 \pm 8.7	11.1 \pm 8.7
ED visits—mean \pm SD	1.9 \pm 3.5	1.8 \pm 3.5	2.1 \pm 3.7	2.0 \pm 3.8	2.3 \pm 4.0	2.2 \pm 4.0	2.4 \pm 4.3	2.4 \pm 4.4
Inpatient visits—mean \pm SD	0.25 \pm 1.0	0.29 \pm 1.0	0.26 \pm 1.0	0.32 \pm 1.1	0.28 \pm 1.1	0.34 \pm 1.2	0.31 \pm 1.1	0.37 \pm 1.3
Hospitalized days—mean \pm SD	1.6 \pm 8.0	1.9 \pm 7.8	1.7 \pm 8.3	2.1 \pm 8.7	1.9 \pm 8.6	2.3 \pm 9.9	2.1 \pm 9.5	2.4 \pm 10.5

Bolded values indicate statistically significant difference ($p < 0.05$) comparing groups 2 and 3 within the month.

received a 30-day or less supply of HCQ, the small subset receiving 60- or 90-day supplies may appear adherent for longer than they were. We conducted a sensitivity analyses looking specifically at this group and found similar trajectory patterns but the declines in adherence, as would be expected, started 2–3 months later than in our primary model. The mean \pm SD PDC for the 91% with 30-day or less HCQ prescriptions was $41\% \pm 29\%$, very much the same as our full cohort. We also used a monthly cutoff of $\geq 80\%$ (24 of 30 days covered) to classify a patient as adherent in keeping with prior chronic disease studies. However, it is unclear if there is a difference in clinical outcomes associated with this cutoff. In addition, while we feel that medically indicated discontinuation within the first year of use of HCQ is uncommon, it is certainly possible that some patients may have stopped their medication because they were told to do so by their physician and we cannot distinguish this from nonadherence using claims data. Similarly, the claims-based definition we used to identify SLE patients may have identified individuals with “probable SLE” as well, as these early undifferentiated patients are often prescribed HCQ. However, as HCQ is most often the initial medication prescribed for SLE, and because retinal toxicity is rare early on, we aimed to understand adherence in the first year of use. Side effects such as gastrointestinal upset, allergic reaction, or hyperpigmentation [46], may occasionally preclude its continued use. While HCQ has been shown to be safe during pregnancy and breastfeeding, it is possible that some women may have chosen to discontinue after consultation with their physicians. This may be reflected by the increased odds of belonging to a nonadherent trajectory among younger age groups. In addition, the use of claims data limits our ability to understand SLE disease activity, which may parallel adherence patterns. We did account for predictors that are markers of SLE disease activity and severity such as the SLE risk adjustment index, medication use (corticosteroid dose and immunosuppressive medications), and comorbidities such as lupus nephritis. We also do not have information regarding disease duration since this is not an inception cohort of SLE patients. In addition, while we were able to examine healthcare utilization (number of hospitalizations, outpatient and emergency department visits), we do not have data available regarding continuity of outpatient care, or access to subspecialty care, which may also influence adherence behavior.

The Medicaid population is a high-risk, vulnerable population with a high burden of comorbidities and adverse outcomes. Therefore, while it is an important population to study, findings might not be broadly generalizable as nonadherence is likely higher in this population. It is possible that there may be misclassification, particularly of comorbidities, since primarily ICD-9 codes alone were used for identification. It is also possible that there are important predictors, such as individual-level socio-economic status, which are not available in Medicaid claims data but may play a significant role in adherence behavior. We also restricted our population to adults because there may be different factors associated with adherence among children and adolescents with SLE. Further studies are needed to examine adherence behavior among adolescents, particularly during the vulnerable transition from pediatric to adult care.

This work also has a number of strengths. We included a large non-academic cohort of HCQ initiators with SLE and used a well-established method not previously applied to SLE medication use that enabled us to understand patterns of adherence over the first year of use. Our patient population was racially and ethnically diverse and 28 U.S. states were represented and we adjusted our analyses by state to account for potential differences in Medicaid enrollment and drug policies. Adherence was measured longitudinally rather than cross-sectionally as it has been done in most prior SLE-related studies, and therefore we captured changes in adherence over time. We restricted our cohort to HCQ initiators

and therefore did not conflate patterns and risk factors among patients receiving this medication for years with those for whom it was newly prescribed. In addition, rather than using a long-term average measure alone, such as the PDC, as most prior claims-based studies do, we used month-by-month measures which have been shown to better represent the nuances of adherence behaviors [13]. While we found a similar percentage formed the persistently adherent trajectory (17%) as were classified as adherent using the PDC composite measure (15%), we were able to delve into predictors of different patterns of nonadherence. Notably, we were able to understand certain factors that contributed to individuals trending from intermediate to complete nonadherence, which has the potential to inform strategies physicians take to counsel patients and identify the most vulnerable groups.

Overall, our study demonstrated that HCQ adherence is a dynamic behavior that declines over the first year of use. While claims data do not allow us to understand the reasons for non-adherence, it is clear from our findings that the majority of patients prescribed HCQ are not taking the medication as prescribed often beginning the month after first dispensing. Potentially modifiable factors, such as improving sustained access to healthcare not only for those who are more severely ill, might prevent intermediate nonadherers, or “undecided” patients from moving toward increasingly nonadherent pathways. In addition, with the knowledge of the extremely poor adherence among this especially vulnerable patient population, increased counseling and support programs both at the time of first HCQ prescription and throughout the first year of use, are needed to promote more sustained patterns of adherence for all patients.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at <https://doi.org/10.1016/j.semarthrit.2018.01.002>.

References

- [1] Marengo MF, Wainmann CA, de Achaval S, Zhang H, Garcia-Gonzalez A, Richardson MN, et al. Measuring therapeutic adherence in systemic lupus erythematosus with electronic monitoring. *Lupus* 2012;21(11):1158–65.
- [2] The Canadian Hydroxychloroquine Study Group. A randomized study of the effect of withdrawing hydroxychloroquine sulfate in systemic lupus erythematosus. *N Engl J Med* 1991;324(3):150–4.
- [3] Akhavan PS, Su J, Lou W, Gladman DD, Urowitz MB, Fortin PR. The early protective effect of hydroxychloroquine on the risk of cumulative damage in patients with systemic lupus erythematosus. *J Rheumatol* 2013;40(6):831–41.
- [4] Pons-Estel GJ, Alarcon GS, Gonzalez LA, Zhang J, Vila LM, Reveille JD, et al. Possible protective effect of hydroxychloroquine on delaying the occurrence of integument damage in lupus: LXI, data from a multiethnic cohort. *Arthritis Care Res (Hoboken)* 2010;62(3):393–400.
- [5] Durcan L, Clarke WA, Magder LS, Petri M. Hydroxychloroquine blood levels in systemic lupus erythematosus: clarifying dosing controversies and improving adherence. *J Rheumatol* 2015;42(11):2092–7.
- [6] Fessler BJ, Alarcon GS, McGwin G Jr, Roseman J, Bastian HM, Friedman AW, et al. Systemic lupus erythematosus in three ethnic groups: XVI. Association of hydroxychloroquine use with reduced risk of damage accrual. *Arthritis Rheum* 2005;52(5):1473–80.
- [7] Ruiz-Itzistorua G, Ramos-Casals M, Brito-Zeron P, Khamashta MA. Clinical efficacy and side effects of antimalarials in systemic lupus erythematosus: a systematic review. *Ann Rheum Dis* 2010;69(1):20–8.
- [8] Melles RB, Marmor MF. The risk of toxic retinopathy in patients on long-term hydroxychloroquine therapy. *JAMA Ophthalmol* 2014;132(12):1453–60.
- [9] Mushlin AI, Appel FA. Diagnosing potential noncompliance. Physicians' ability in a behavioral dimension of medical care. *Arch Intern Med* 1977;137(3):318–21.
- [10] Koneru S, Shishov M, Ware A, Farhey Y, Mongey AB, Graham TB, et al. Effectively measuring adherence to medications for systemic lupus erythematosus in a clinical setting. *Arthritis Rheum* 2007;57(6):1000–6.
- [11] Nagin DS, Odgers CL. Group-based trajectory modeling (nearly) two decades later. *J Quant Criminol* 2010;26(4):445–53.

- [12] Nagin DS, Odgers CL. Group-based trajectory modeling in clinical research. *Annu Rev Clin Psychol* 2010;6:109–38.
- [13] Franklin JM, Shrank WH, Pakes J, Sanfelix-Gimeno G, Matlin OS, Brennan TA, et al. Group-based trajectory models: a new approach to classifying and predicting long-term medication adherence. *Med Care* 2013;51(9):789–96.
- [14] Li Y, Zhou H, Cai B, Kahler KH, Tian H, Gabriel S, et al. Group-based trajectory modeling to assess adherence to biologics among patients with psoriasis. *Clinicoecon Outcomes Res* 2014;6:197–208.
- [15] Juarez DT, Williams AE, Chen C, Daida YG, Tanaka SK, Trinacty CM, et al. Factors affecting medication adherence trajectories for patients with heart failure. *Am J Manag Care* 2015;21(3):e197–205.
- [16] MacEwan JP, Forma FM, Shafrazi J, Hatch A, Lakdawalla DN, Lindenmayer JP. Patterns of adherence to oral atypical antipsychotics among patients diagnosed with schizophrenia. *J Manag Care Spec Pharm* 2016;22(11):1349–61.
- [17] Feldman CH, Hiraki LT, Liu J, Fischer MA, Solomon DH, Alarcon GS, et al. Epidemiology and sociodemographics of systemic lupus erythematosus and lupus nephritis among US adults with Medicaid coverage, 2000–2004. *Arthritis Rheum* 2013;65(3):753–63.
- [18] Kim SY, Servi A, Polinski JM, Mogun H, Weinblatt ME, Katz JN, et al. Validation of rheumatoid arthritis diagnoses in health care utilization data. *Arthritis Res Ther* 2011;13(1):R32.
- [19] Barnado A, Casey C, Carroll RJ, Wheless L, Denny JC, Crofford LJ. Developing electronic health record algorithms that accurately identify patients with systemic lupus erythematosus. *Arthritis Care Res (Hoboken)* 2017;69(5):687–93.
- [20] Farhangian ME, Huang WW, Feldman SR. Adherence to oral and topical medications in cutaneous lupus erythematosus is not well characterized. *Dermatol Ther (Heidelberg)* 2015;5(2):91–105.
- [21] Gomez-Puerta JA, Barboura M, Guan H, Feldman CH, Alarcon GS, Costenbader KH. Racial/ethnic variation in all-cause mortality among United States Medicaid recipients with systemic lupus erythematosus: a Hispanic and Asian paradox. *Arthritis Rheumatol* 2015;67(3):752–60.
- [22] Hudson M, Suissa S. Avoiding common pitfalls in the analysis of observational studies of new treatments for rheumatoid arthritis. *Arthritis Care Res (Hoboken)* 2010;62(6):805–10.
- [23] Steiner JF, Prochazka AV. The assessment of refill compliance using pharmacy records: methods, validity, and applications. *J Clin Epidemiol* 1997;50(1):105–16.
- [24] Steiner JF, Koepsell TD, Fihn SD, Inui TS. A general method of compliance assessment using centralized pharmacy records. Description and validation. *Med Care* 1988;26(8):814–23.
- [25] Grymonpre R, Cheang M, Fraser M, Metge C, Sitar DS. Validity of a prescription claims database to estimate medication adherence in older persons. *Med Care* 2006;44(5):471–7.
- [26] Nacheva JB, Hislop M, Dowdy DW, Lo M, Omer SB, Regensburg L, et al. Adherence to highly active antiretroviral therapy assessed by pharmacy claims predicts survival in HIV-infected South African adults. *J Acquir Immune Defic Syndr* 2006;43(1):78–84.
- [27] Pladenvall M, Williams LK, Potts LA, Divine G, Xi H, Lafata JE. Clinical outcomes and adherence to medications measured by claims data in patients with diabetes. *Diabetes Care* 2004;27(12):2800–5.
- [28] Steiner JF, Fihn SD, Blair B, Inui TS. Appropriate reductions in compliance among well-controlled hypertensive patients. *J Clin Epidemiol* 1991;44(12):1361–71.
- [29] Yeaw J, Benner JS, Walt JG, Sian S, Smith DB. Comparing adherence and persistence across 6 chronic medication classes. *J Manag Care Pharm* 2009;15(9):728–40.
- [30] Manson S, Schroeder J, Van Riper D, Ruggles S. IPUMS National Historical Geographic Information System: Version 12.0 [database]. Minneapolis: University of Minnesota. 2017. <http://dx.doi.org/http://doi.org/10.18128/0050.v12.0>.
- [31] Ward MM. Development and testing of a systemic lupus-specific risk adjustment index for in-hospital mortality. *J Rheumatol* 2000;27(6):1408–13.
- [32] Chibnik LB, Massarotti EM, Costenbader KH. Identification and validation of lupus nephritis cases using administrative data. *Lupus* 2010;19(6):741–3.
- [33] Noyes K, Liu H, Lyness JM, Friedman B. Medicare beneficiaries with depression: comparing diagnoses in claims data with the results of screening. *Psychiatr Serv* 2011;62(10):1159–66.
- [34] Twisk J, Hoekstra T. Classifying developmental trajectories over time should be done with great caution: a comparison between methods. *J Clin Epidemiol* 2012;65(10):1078–87.
- [35] Nagin DS. Group-Based Modeling of Development. Boston, MA: Harvard University Press; 2005.
- [36] Feldman CH, Yazdany J, Guan H, Solomon DH, Costenbader KH. Medication nonadherence is associated with increased subsequent acute care utilization among Medicaid beneficiaries with systemic lupus erythematosus. *Arthritis Care Res (Hoboken)* 2015;67(12):1712–21.
- [37] Oliveira-Santos M, Verani JF, Klumb EM, Albuquerque EM. Evaluation of adherence to drug treatment in patients with systemic lupus erythematosus in Brazil. *Lupus* 2011;20(3):320–9.
- [38] Jung H, Bobba R, Su J, Shariati-Sarabi Z, Gladman DD, Urowitz M, et al. The protective effect of antimarial drugs on thrombovascular events in systemic lupus erythematosus. *Arthritis Rheum* 2010;62(3):863–8.
- [39] Julian LJ, Yelin E, Yazdany J, Panopalis P, Trupin L, Criswell LA, et al. Depression, medication adherence, and service utilization in systemic lupus erythematosus. *Arthritis Rheum* 2009;61(2):240–6.
- [40] Feldman CH, Collins JE, Zhang Z, Kawachi I, Solomon DH, Costenbader KH. Azathioprine and mycophenolate mofetil adherence in a nationwide Medicaid cohort with systemic lupus erythematosus [abstract]. *Arthritis Rheumatol*. 2016;2016(68):Suppl 10.
- [41] Mehat P, Atiquzzaman M, Esdaile JM, AviNa-Zubieta A, De Vera MA. Medication nonadherence in systemic lupus erythematosus: a systematic review. *Arthritis Care Res (Hoboken)* 2017;69(11):1706–13.
- [42] Choudhry NK, Krumme AA, Ercole PM, Girdish C, Tong AY, Khan NF, et al. Effect of reminder devices on medication adherence: the REMIND Randomized Clinical Trial. *JAMA Intern Med* 2017;177(5):624–31.
- [43] Bailey JE, Hajjar M, Shoib B, Tang J, Ray MM, Wan JY. Risk factors associated with antihypertensive medication nonadherence in a statewide Medicaid population. *Am J Med Sci* 2014;348(5):410–5.
- [44] Garcia-Gonzalez A, Richardson M, Garcia Popa-Lisseanu M, Cox V, Kallen MA, Janssen N, et al. Treatment adherence in patients with rheumatoid arthritis and systemic lupus erythematosus. *Clin Rheumatol* 2008;27(7):883–9.
- [45] Mosley-Williams A, Lumley MA, Gillis M, Leisen J, Guice D. Barriers to treatment adherence among African American and white women with systemic lupus erythematosus. *Arthritis Rheum* 2002;47(6):630–8.
- [46] Jallouli M, Frances C, Piette JC, Huong du LT, Moguelet P, Factor C, et al. Hydroxychloroquine-induced pigmentation in patients with systemic lupus erythematosus: a case-control study. *JAMA Dermatol* 2013;149(8):935–40.

Article type : Brief Report

Azathioprine and Mycophenolate Mofetil Adherence Patterns and Predictors among Medicaid Beneficiaries with Systemic Lupus Erythematosus

Candace H. Feldman, MD, ScD¹, Jamie Collins, PhD², Zhi Zhang, MS¹, Chang Xu¹, SV Subramanian, PhD³, Ichiro Kawachi, MB.ChB, PhD³, Daniel H. Solomon, MD, MPH¹, Karen H. Costenbader, MD, MPH¹

¹ Division of Rheumatology, Immunology and Allergy, Department of Medicine, Brigham and Women's Hospital, Boston, MA

² The Orthopaedic and Arthritis Center for Outcomes Research (OrACORe), Brigham and Women's Hospital, Boston, MA

³ Department of Social Behavioral Sciences, Harvard T. H. Chan School of Public Health, Boston MA

Corresponding author:

Candace H. Feldman, MD, ScD
60 Fenwood Road, Office 6016P
Boston, MA 02115
Phone: 617-525-1035
Fax: 617-264-3019
Email: cfeldman@bwh.harvard.edu

Keywords: Adherence, azathioprine, mycophenolate mofetil, systemic lupus erythematosus, pharmacoepidemiology, observational study, health services research

Funding: This study was funded by the Rheumatology Research Foundation Investigator Award (CH Feldman), NIH K23 AR071500 (CH Feldman), R01 057327 (KH Costenbader), K24 AR066109 (KH Costenbader), K24 AR055989 (DH Solomon) and P30 AR072577 (DH Solomon and KH Costenbader).

None of the authors have relevant financial disclosures.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/acr.23792

This article is protected by copyright. All rights reserved.

Abstract

Objective: Azathioprine (AZA) and mycophenolate mofetil (MMF) are frequently used immunosuppressives for moderate-to-severe SLE. We studied longitudinal patterns and predictors of adherence to AZA and MMF in a nationwide U.S. SLE cohort.

Methods: In the Medicaid Analytic eXtract (2000-2010), we identified SLE patients who initiated AZA or MMF (no use in prior 6 months) with ≥ 12 months of continuous follow-up. We dichotomized adherence at 80% with $\geq 24/30$ days/month considered adherent. We used group-based trajectory models to estimate monthly adherence patterns and multivariable multinomial logistic regression to determine the association between demographic, SLE and utilization-related predictors and the odds ratios (OR) of belonging to a nonadherent vs. the adherent trajectory, separately for AZA and MMF.

Results: We identified 2,309 AZA initiators and 2,070 MMF initiators with SLE. Four-group trajectory models classified 17% of AZA and 21% of MMF initiators as adherent. AZA and MMF nonadherers followed similar trajectory patterns. Black race (OR 1.67, 95% CI 1.20-2.31) and Hispanic ethnicity (OR 1.58, 95% CI 1.06-2.35) increased odds of AZA nonadherence; there were no significant associations between race/ethnicity and MMF nonadherence. Male sex and polypharmacy were associated with lower odds of nonadherence to both medications; lupus nephritis was associated with lower odds of nonadherence to MMF (OR 0.74, 95% CI 0.55-0.99).

Conclusions: Adherence to AZA or MMF over the first year of use was rare. Race, sex and lupus nephritis were modestly associated with adherence, but the magnitude, direction and significance of predictors differed by medication suggesting the complexity of predicting adherence behavior.

Significance and Innovation

- Adherence to the two most frequently used immunosuppressive medications for SLE, azathioprine and mycophenolate, is overall extremely poor over the first year of use among a national cohort of Medicaid beneficiaries with SLE.
- Adherence overall was slightly better to mycophenolate mofetil compared to azathioprine however adherence to both medications declined significantly for nearly 80 percent of patients over the first year of use.
- While demographic factors including black race, Hispanic ethnicity and younger age were associated with higher odds of nonadherence among azathioprine initiators, they were significantly less strongly associated with nonadherence among mycophenolate mofetil initiators. This suggests that a single set of patient characteristics does not consistently predict nonadherence patterns across medications.

Background

Systemic lupus erythematosus (SLE) is a complex autoimmune disease with a range of organ system manifestations. Patients with moderate-to-severe disease often receive immunosuppressives, either azathioprine (AZA) or mycophenolate mofetil (MMF) often interchangeably, to control lupus nephritis, serositis, hematologic abnormalities, arthritis and cutaneous disease.(1, 2) Adherence to medications for SLE varies from 20-80% depending on the population studied, the medication, and the method used to measure adherence (e.g. self-reported surveys, blood levels, prescription refill data).(3-6) Higher rates of nonadherence have been observed among younger age groups, black and Hispanic patients, and individuals with less education.(4, 7, 8) Studies have varied as to whether polypharmacy and disease severity affect the risk for nonadherence.

In Medicaid, the largest public insurance in the U.S. primarily for low income individuals, our prior work demonstrated that under 20% of patients were adherent to hydroxychloroquine (HCQ), defined as $\geq 80\%$ of days covered with prescription refills during the first year of use.(8) There were higher odds of nonadherence among SLE patients who were young, black, taking fewer medications, and with less severe disease. We also observed that HCQ adherence was dynamic and for most, declined over the first year of use. In this study, we aimed to assess patterns of adherence over the first year of AZA or MMF use. We hypothesized that like HCQ, adherence would decline over time and that while predictors of nonadherence would be similar, patterns would suggest better adherence to MMF because patients may be slightly sicker and therefore more invested in continuing their medication.

Methods

Patient Cohort

We used the Medicaid Analytic eXtract (MAX) with demographic data, billing claims, healthcare utilization, and drug dispensing data from 2000-2010 for the 29 most populated U.S. states (86% of Medicaid beneficiaries nationwide). We excluded all claims from Ohio because detailed medication dispensing data were not available, and additionally excluded all individuals without drug dispensing data, including those who were hospitalized for the entire follow-up period. We identified two cohorts of SLE patients (≥ 2 ICD-9 SLE codes (710.0) for discharge diagnoses or physician claims ≥ 30 days apart) with either AZA or MMF dispensing within 365 days of a SLE code.(8) We required ≥ 6 months of continuous enrollment without use of AZA or MMF prior to the date of initiation (index date). We allowed AZA initiators to previously receive MMF and vice versa. We required ≥ 365 days of continuous enrollment following the index date to assess adherence.

Measures of Adherence

We used prescription refill data to measure adherence, which has previously been validated in claims data.(9) We calculated the proportion of days covered (PDC) beginning at the index date for 365 days (number of days covered divided by 365 times 100, subtracting hospitalized days from numerator and denominator) and classified individuals with PDC \geq 80% as adherent.(10) We also measured adherence monthly to MMF and to AZA over the 12-month period and each month was classified either as adherent (1) or nonadherent (0) depending on whether \geq 24 of 30 days (80%) were covered. The majority ($>85\%$) of our cohort received a one-month supply of their AZA or MMF in accordance with Medicaid policies in most states.

Potential Correlates of Adherence

We measured potential predictors during the 6 months prior to and including the index date in the AZA and MMF cohorts. Demographic factors included age and state of residence at the index date, sex, race/ethnicity and region from MAX, and zip code-level median household income from American Community Survey data.(11) We included diabetes, smoking, lupus nephritis, antidepressant medication use, SLE-related laboratory tests and medications (HCQ, immunosuppressives and corticosteroids), number of medications at the index date, days' supply of first AZA/MMF fill, vaccinations, healthcare utilization, and the SLE risk-adjustment index (**Supplemental Table 1**).^(8, 12, 13) We also examined models that included comorbidities (thromboembolism, pulmonary, hepatic, cardiovascular and cerebrovascular disease, substance abuse, obesity, and malignancy) but did not include them in our final models as they did not contribute significantly and have not been shown in prior studies to be strongly associated with adherence.

Statistical Analyses

We compared baseline characteristics and PDC for AZA and MMF using descriptive statistics. We used our binary indicators of monthly adherence to construct group-based trajectory models (GBTM) to classify patients by adherence, separately for AZA and MMF. GBTM are used to identify latent patterns in longitudinal data with repeated measures and has been previously applied to prescription refill data to uncover adherence patterns over the first year of use.(8, 14) We evaluated AZA and MMF GBTMs ranging from three to six trajectory groups and based our model choice on a combination of Bayesian information criteria with lower values considered preferable, reasonable distribution across groups, posterior probabilities $\geq 80\%$ for each group, and explanatory potential.(15) We then used multinomial logistic regression models for both AZA and MMF to determine the odds of belonging to a nonadherent trajectory compared to the persistently adherent trajectory for demographic, utilization and SLE-related predictors.

We conducted sensitivity analyses censoring at potential conditions that may have resulted in physician-recommended discontinuation rather than nonadherence. We censored at the beginning of the nearest preceding refill for first discharge diagnosis code for serious infection, and for any code for neutropenia or transaminitis for both cohorts, and additionally for pregnancy or colitis for the MMF cohort (**Supplemental Table 1**). We conducted all analyses using SAS 9.4 (Cary, NC) and used the “Proc Traj” add-on package for GBTM. Data were obtained from the Centers for Medicare and Medicaid Services through a Data Use Agreement and data are presented in accordance with their policies. The Partners Healthcare IRB approved this study.

Results

We identified 2,309 AZA initiators and 2,070 MMF initiators. AZA initiators were slightly older with a mean \pm SD age of 36.1 ± 11.8 compared to 33.4 ± 11.6 for MMF (**Table 1**). The percentage of females was slightly higher among AZA initiators, as was the percentage of black individuals. On average, at baseline AZA initiators had less severe SLE with a lower mean SLE risk adjustment index, lower prevalence of lupus nephritis, fewer overall medications, and less immunosuppressant use. Corticosteroid use was comparable between initiators of the two drugs. The mean \pm SD PDC for AZA initiators beginning at the index date of new use was $40\% \pm 29$, with 15% classified as adherent ($PDC \geq 80\%$) compared to $44\% \pm 30\%$ for MMF, with 18% classified as adherent ($p < 0.001$). We did not observe significant changes in either AZA or MMF adherence by index date year (**Supplemental Figure 1**). When we varied the adherence threshold, 7.3% of AZA initiators and 8.6% of MMF initiators had PDCs $\geq 90\%$, and 21.9% of AZA initiators and 25.7% of MMF initiators had PDCs $\geq 70\%$.

We examined three to six-group GBTMs for both AZA and MMF initiators. We aimed to balance model fit with explanatory power and to compare similar numbers of trajectories for both medications. A four-group model provided an adequate fit for both drugs. The mean posterior probabilities for each trajectory were greater than 80%, and each trajectory had a reasonably balanced distribution of individuals. The Bayesian information criteria for the three-group models were slightly smaller than for the four-group models but we chose the four-group models because we felt that the explanatory potential was greater and the other model fit criteria were met.(15)

Overall, we observed similar patterns for the four-group trajectory model (**Figure 1**). In the persistently adherent trajectory (group 4, in black), there were 384 (17%) AZA initiators and 441 (21%) MMF initiators. In the persistently nonadherent trajectory (group 1,

in red), there were 1030 (45%) AZA initiators and 883 (43%) MMF initiators. Among AZA initiators, two groups with more dynamic nonadherent patterns (group 2, in green and group 3, in blue), steadily declined until between 6-7 months at which point adherence for group 3 plateaued at about 45% of days/month covered, while group 2 continued to decline. Among MMF initiators, among those with more dynamic nonadherent patterns, group 2 (in green) precipitously declined initially and then plateaued at about 30% of days/month covered. Group 3 (in blue) declined more slowly over the course of use, remaining just below the adherent range of $\geq 80\%$ of days covered until between months 4-5.

We examined multinomial logistic regression models comparing the odds of belonging to the nonadherent trajectories (groups 1-3) vs. the persistently adherent trajectory (group 4) for both AZA and MMF initiators (**Table 2**). Among AZA initiators, we observed increased odds of belonging to the persistently nonadherent trajectory (group 1) vs. the persistently adherent (group 4) among SLE patients who were black (OR 1.67, 95% CI 1.20-2.31) or Hispanic (OR 1.58, 95% CI 1.06-2.35) compared to White, and in the 18-35-year age group (OR 1.60, 95% CI 1.10-2.34) compared to the oldest (age 51-65 years). We found greater than two times higher odds of belonging to the declining and then plateauing nonadherent trajectory (group 3), vs. the persistently adherent trajectory (group 4) for black race and Hispanic ethnicity, and reduced odds of belonging to group 3 vs. 4 among individuals living in areas with less than or equal to the median of the zip code median household income, compared to areas above the median (OR 0.63, 95% CI 0.47-0.86). Male sex and increased number of medications were associated with reduced odds of belonging to nearly all nonadherent trajectories compared to the most adherent (group 4).

Among MMF initiators, we did not observe statistically significant associations by race/ethnicity comparing any of the nonadherent trajectories (groups 1-3) to the persistently adherent trajectory (group 4). Like AZA, we observed reduced odds of belonging to the

persistently nonadherent trajectory (group 1) vs. the persistently adherent (group 4) among males compared to females (OR 0.67, 95% CI 0.45-0.99) and for increased medication use (OR 0.90, 95% CI 0.87-0.94). Specific to MMF initiators, we observed reduced odds of belonging to the persistently nonadherent vs. the persistently adherent group among SLE patients with lupus nephritis (OR 0.74, 95% CI 0.55-0.99) and increased odds among those living in areas below or equal to the median household income compared to above (OR 1.33, 95% CI 1.02-1.72). Increased number of emergency department visits were also associated with increased odds of belonging to least adherent groups (1 and 2) vs. the most adherent group (4).

In sensitivity analyses, censoring at indications that may have resulted in physician recommended discontinuation of AZA or MMF, we observed only modestly increased adherence estimates. Among AZA initiators, censoring at ≥ 1 ICD-9 code for serious infection, transaminitis or neutropenia, the mean \pm SD PDC was $45\% \pm 30$ with 18.8% categorized as adherent (PDC $\geq 80\%$). Among MMF initiators, censoring at serious infection, transaminitis, neutropenia, colitis or pregnancy, the mean (SD) PDC was $49\% \pm 29$ with 22.7% categorized as adherent.

Discussion

Overall, we observed profoundly poor adherence among Medicaid beneficiaries with SLE who initiated AZA or MMF over the first year of use; less than a quarter of patients refilled their medications 80% of the time or more. These rates of nonadherence were similar among HCQ initiators also enrolled in Medicaid.(8) Our findings were also in line with a prior study that utilized trajectory models to describe adherence patterns among statin initiators; only 23.4% were persistently adherent over the first 15 months of use.(14)

In our cohorts, adherence was slightly better among MMF initiators compared to AZA initiators, and groups of MMF nonadherers appeared to remain at least partially adherent for longer. The populations of AZA and MMF initiators were somewhat different – AZA initiators were slightly older, included more females, likely because this medication is compatible with pregnancy whereas MMF is not, and had less severe SLE, with a lower prevalence of lupus nephritis and prior immunosuppressive use. Predictors of nonadherence differed as well. We observed primarily demographic associations with AZA nonadherence; black race, Hispanic ethnicity, female sex, and younger age were associated with increased odds of nonadherence. However, among MMF initiators, while we did see an association between female sex, younger age and nonadherence, the findings were less consistent across nonadherent trajectories, and we did not observe significant associations by race/ethnicity.

Among MMF initiators, we observed reduced odds of persistent nonadherence associated with lupus nephritis and increased medication use and increased odds associated with visiting the emergency room more. The median zip-code level household income was slightly higher for MMF initiators compared to AZA, and interestingly, while we observed increased odds of nonadherence among MMF initiators living in lower vs. higher median income zip codes, the trend was in the opposite direction among AZA initiators for unclear reasons.

In accordance with prior studies, we did not find one dominant factor to be consistently associated with all nonadherence patterns across different SLE-related medications. While many factors likely contribute to adherence behavior, their relationship is not necessarily constant across medications, populations, or over time of use. Differences between SLE-related medications including affordability, tolerability, regimen complexity, interactions with other medications or lifestyle factors, likely contribute to varying patterns of adherence. In addition, aspects that cannot be measured in a study like ours that relies on claims data including the doctor-patient relationship, or beliefs about SLE and about

medication safety, play a role as well. This complexity suggests that a simple variable-based algorithm to predict a person's adherence pattern over time to multiple medications is likely unrealistic.

This study has limitations. We utilized prescription refill data to determine adherence, which may not always represent use, however prior studies have shown this to be a valid method.(9, 16) We did not have data regarding initial AZA or MMF prescribing and therefore could not capture primary nonadherence. We used a cutoff of $\geq 80\%$ to indicate adherence, and while this is accepted in the chronic disease literature, it is unclear whether it correlates with the physiologic levels needed to have a clinically meaningful effect. In addition, we could not distinguish between medically-indicated discontinuation and nonadherence. However, we excluded all hospitalized time during which adherence could not be readily measured and conducted sensitivity analyses censoring at potential indications for physician recommended discontinuation which resulted in adherence estimates similar to our primary analyses. We lacked qualitative measures of potential predictors of nonadherence, as well as actual lab results and medical records to understand fluctuations in disease activity. We also lacked information on socioeconomic status, such as individual income or education, and we cannot exclude the possibility that racial/ethnic differences may be markers for differences in socioeconomic circumstances. We lacked more recent data past 2010, however between 2000-2010, we did not appreciate significant fluctuations in either AZA or MMF adherence.

In our study, we leveraged two large cohorts of AZA and MMF initiators to examine adherence patterns over time. While we found relatively similar patterns between the two medications, MMF initiators seemed to stay at least partially adherent for longer, possibly due to an understanding of the need to treat more severe manifestations such as lupus nephritis. Adherence overall however, was very poor and in this vulnerable, low-income SLE

population, more needs to be done to facilitate persistent adherence to efficacious, standard-of-care medications to ultimately reduce disparities in outcomes.

Figure 1. Group-based trajectory models demonstrating monthly adherence patterns for SLE patients enrolled in Medicaid over the first year of azathioprine and mycophenolate mofetil use, with group 4 (in black) as the persistently adherent trajectory and group 1 (in red) as the persistently nonadherent trajectory

References

1. Hahn BH, McMahon MA, Wilkinson A, Wallace WD, Daikh DI, Fitzgerald JD, et al. American College of Rheumatology guidelines for screening, treatment, and management of lupus nephritis. *Arthritis Care Res (Hoboken)*. 2012;64(6):797-808.
2. Tunnicliffe DJ, Singh-Grewal D, Kim S, Craig JC, Tong A. Diagnosis, Monitoring, and Treatment of Systemic Lupus Erythematosus: A Systematic Review of Clinical Practice Guidelines. *Arthritis Care Res (Hoboken)*. 2015;67(10):1440-52.
3. Julian LJ, Yelin E, Yazdany J, Panopalis P, Trupin L, Criswell LA, et al. Depression, medication adherence, and service utilization in systemic lupus erythematosus. *Arthritis Rheum*. 2009;61(2):240-6.
4. Durcan L, Clarke WA, Magder LS, Petri M. Hydroxychloroquine Blood Levels in Systemic Lupus Erythematosus: Clarifying Dosing Controversies and Improving Adherence. *J Rheumatol*. 2015;42(11):2092-7.
5. Feldman CH, Yazdany J, Guan H, Solomon DH, Costenbader KH. Medication Nonadherence Is Associated With Increased Subsequent Acute Care Utilization Among Medicaid Beneficiaries With Systemic Lupus Erythematosus. *Arthritis Care Res (Hoboken)*. 2015;67(12):1712-21.
6. Costedoat-Chalumeau N, Houssiau F, Izmirly P, Guern VL, Navarra S, Jolly M, et al. A Prospective International Study on Adherence to Treatment in 305 Patients With Flaring SLE: Assessment by Drug Levels and Self-Administered Questionnaires. *Clin Pharmacol Ther*. 2018.
7. Garcia-Gonzalez A, Richardson M, Garcia Popa-Lisseanu M, Cox V, Kallen MA, Janssen N, et al. Treatment adherence in patients with rheumatoid arthritis and systemic lupus erythematosus. *Clin Rheumatol*. 2008;27(7):883-9.
8. Feldman CH, Collins J, Zhang Z, Subramanian SV, Solomon DH, Kawachi I, et al. Dynamic patterns and predictors of hydroxychloroquine nonadherence among Medicaid beneficiaries with systemic lupus erythematosus. *Semin Arthritis Rheum*. 2018.
9. Steiner JF, Prochazka AV. The assessment of refill compliance using pharmacy records: methods, validity, and applications. *Journal of clinical epidemiology*. 1997;50(1):105-16.
10. Yeaw J, Benner JS, Walt JG, Sian S, Smith DB. Comparing adherence and persistence across 6 chronic medication classes. *J Manag Care Pharm*. 2009;15(9):728-40.
11. Manson S, Schroeder J, Van Riper D, Ruggles S. IPUMS National Historical Geographic Information System, Version 12.0 [Database]. Minneapolis: University of Minnesota; 2017.
12. Ward MM. Development and testing of a systemic lupus-specific risk adjustment index for in-hospital mortality. *J Rheumatol*. 2000;27(6):1408-13.
13. Chibnik LB, Massarotti EM, Costenbader KH. Identification and validation of lupus nephritis cases using administrative data. *Lupus*. 2010;19(6):741-3.
14. Franklin JM, Shrank WH, Pakes J, Sanfelix-Gimeno G, Matlin OS, Brennan TA, et al. Group-based trajectory models: a new approach to classifying and predicting long-term medication adherence. *Med Care*. 2013;51(9):789-96.

15. Nagin DS. Group-based modeling of development. Cambridge, MA: Harvard University Press; 2005.

16. Steiner JF, Koepsell TD, Fihn SD, Inui TS. A general method of compliance assessment using centralized pharmacy records. Description and validation. Medical care. 1988;26(8):814-23.

Table 1. Baseline characteristics of azathioprine and mycophenolate initiators with SLE in Medicaid, 2000-2010

Characteristics	Azathioprine (N=2309)	Mycophenolate mofetil (N=2070)
Age – mean \pm SD	36.1 \pm 11.8	33.4 \pm 11.6
Female sex – N (%)	2138 (92.6)	1857 (89.7)
Race/Ethnicity		
Black	1089 (47.2)	925 (44.7)
White	579 (25.1)	506 (24.4)
Hispanic	476 (20.6)	453 (21.9)
Asian	93 (4.0)	114 (5.5)
American Indian/Alaska Native	22 (1.0)	31 (1.5)
Region – N (%)		
Northeast	523 (22.7)	577 (27.9)
South	837 (36.3)	680 (32.9)
Midwest	378 (16.4)	368 (17.8)
West	571 (24.7)	445 (21.5)
Zip code median household income- median \$ (25th, 75th)	41,643 (33,659-51,948)	42,557 (33,995-55,565)
Diabetes- N (%)	297 (12.9)	242 (11.7)
SLE risk adjustment index – mean \pm SD	1.4 \pm 2.2	2.1 \pm 2.5
Lupus nephritis – N (%)	503 (21.8)	1162 (56.1)
Number of drugs – mean \pm SD	5.0 \pm 3.6	5.5 \pm 3.8
Antidepressant use – N (%)	439 (19.0)	209 (10.1)
Corticosteroid use – N (%)	1898 (82.2)	1704 (82.3)
Hydroxychloroquine use – N (%)	1315 (57.0)	938 (45.3)
Immunosuppressant use* – N (%)	374 (16.2)	526 (25.4)
Number of SLE-related laboratory tests[†] – mean (SD)	3.3 (4.2)	3.71 (5.0)

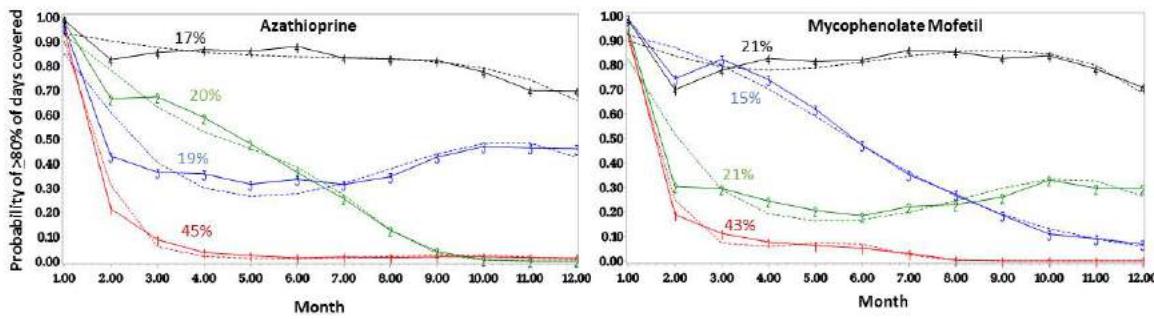
*Immunosuppressants include methotrexate, leflunomide, tacrolimus, sulfasalazine, cyclosporine, cyclophosphamide, azathioprine (for mycophenolate mofetil initiators) and mycophenolate mofetil (for azathioprine initiators)

[†]Laboratory tests include BUN, creatinine, complement (C3 and C4), ESR, CRP, anti-dsDNA and urinalysis

Table 2. Multivariable multinomial regression comparing the odds of belonging to a nonadherent trajectory (groups 1-3) to the most adherent trajectory (group 4, reference) for azathioprine and mycophenolate mofetil initiators with SLE

Predictors	Azathioprine (N=2309)			Mycophenolate Mofetil (N=2070)				
	OR (95% CI)	Group 1	Group 2	Group 3	OR (95% CI)	Group 1	Group 2	Group 3
N (%)	1030 (44.6)	459 (19.9)	436 (18.9)	883 (42.7)	441 (21.3)	305 (14.7)		
Age (ref=51-65)								
18-34 years	1.60 (1.10-2.34)	1.75 (1.12-2.73)	1.48 (0.96-2.29)	1.14 (0.76-1.71)	1.15 (0.71-1.86)	1.95 (1.12-3.42)		
35-50 years	1.36 (0.95-1.96)	1.67 (1.09-2.57)	1.17 (0.76-1.79)	0.83 (0.55-1.26)	0.96 (0.59-1.56)	1.44 (0.82-2.54)		
Male sex (ref=female)	0.59 (0.39-0.91)	0.68 (0.42-1.11)	0.54 (0.32-0.92)	0.67 (0.45-0.99)	0.85 (0.56-1.31)	0.83 (0.51-1.34)		
Race/ethnicity (ref=White)								
Black	1.67 (1.20-2.31)	1.62 (1.11-2.35)	2.05 (1.39-3.02)	1.33 (0.95-1.86)	1.20 (0.82-1.76)	0.72 (0.47-1.09)		
Hispanic	1.58 (1.06-2.35)	1.37 (0.87-2.15)	2.00 (1.26-3.19)	0.94 (0.64-1.39)	0.85 (0.55-1.32)	0.88 (0.56-1.40)		
Asian	1.02 (0.52-1.99)	0.85 (0.40-1.84)	1.52 (0.72-3.18)	0.64 (0.36-1.13)	0.62 (0.33-1.16)	0.59 (0.30-1.17)		
American Indian/Alaska Native	0.78 (0.24-2.51)	0.61 (0.15-2.52)	NR	1.88 (0.57-6.25)	1.49 (0.39-5.65)	0.77 (0.16-3.83)		
SLE risk	0.98 (0.91-1.05)	0.97 (0.90-1.05)	0.98 (0.90-1.06)	0.99 (0.92-1.05)	0.97 (0.90-1.04)	1.00 (0.93-1.08)		

adjustment index						
Lupus nephritis	1.06 (0.75-1.49)	1.22 (0.83-1.79)	1.31 (0.89-1.93)	0.74 (0.55-0.99)	0.98 (0.70-1.36)	0.95 (0.66-1.37)
Diabetes mellitus	1.11 (0.74-1.67)	1.74 (1.11-2.71)	1.30 (0.81-2.08)	1.05 (0.70-1.59)	0.95 (0.59-1.53)	1.02 (0.61-1.68)
Median household income \leq median (ref= >median)	0.80 (0.61-1.04)	0.99 (0.73-1.34)	0.63 (0.47-0.86)	1.33 (1.02-1.72)	1.28 (0.95-1.72)	1.26 (0.92-1.75)
Number of medications	0.90 (0.86-0.93)	0.94 (0.90-0.98)	0.92 (0.88-0.96)	0.90 (0.87-0.94)	0.90 (0.86-0.94)	0.99 (0.95-1.03)
Outpatient visits	1.01 (0.98-1.03)	0.99 (0.97-1.02)	0.99 (0.9-1.01)	0.99 (0.96-1.01)	0.99 (0.96-1.01)	0.98 (0.95-1.00)
Hospitalizations	1.08 (0.96-1.21)	1.05 (0.93-1.20)	1.09 (0.96-1.24)	0.98 (0.89-1.08)	0.96 (0.86-1.08)	0.98 (0.87-1.10)
Emergency Department Visits	1.05 (0.99-1.12)	1.05 (0.98-1.13)	1.01 (0.94-1.09)	1.16 (1.08-1.25)	1.14 (1.05-1.23)	1.06 (0.97-1.15)
Models additionally adjusted for calendar year of index date, state of residence at index date, days' supply at first dispensing, SLE-related medication use (hydroxychloroquine, immunosuppressants, corticosteroids), number of SLE-related laboratory tests, antidepressant use, smoking, obesity, and influenza and pneumococcal vaccinations						





HHS Public Access

Author manuscript

Int J Clin Rheumtol. Author manuscript; available in PMC 2018 March 29.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Published in final edited form as:

Int J Clin Rheumtol. 2018 ; 13(1): 1–10. doi:10.4172/1758-4272.1000154.

Association of weight loss with improved disease activity in patients with rheumatoid arthritis: A retrospective analysis using electronic medical record data

David J. Kreps, MS¹, Florencia Halperin, MD, MMSc^{2,3}, Sonali P. Desai, MD, MPH^{1,2}, Zhi Z. Zhang, MS¹, Elena Losina, PhD^{2,4}, Amber T. Olson¹, Elizabeth W. Karlson, MD, MS^{1,2}, Bonnie L. Bermas, MD⁵, and Jeffrey A. Sparks, MD, MMSc^{1,2}

¹Department of Medicine, Division of Rheumatology, Immunology and Allergy, Brigham and Women's Hospital, Boston, MA

²Harvard Medical School, Boston, MA

³Center for Weight Management and Metabolic Surgery, Department of Medicine, Division of Endocrinology, Diabetes and Hypertension, Brigham and Women's Hospital, Boston, MA

⁴Orthopedic and Arthritis Center for Outcomes Research, Policy and Innovation evaluation of Orthopedic Treatments (PIVOT) Center, Department of Orthopedic Surgery, Brigham and Women's Hospital, Boston, MA

⁵University of Texas Southwestern Medical Center, Dallas, TX

Abstract

Objective—To evaluate the association between weight loss and rheumatoid arthritis (RA) disease activity.

Methods—We conducted a retrospective cohort study of RA patients seen at routine clinic visits at an academic medical center, 2012–2015. We included patients who had 2 clinical disease activity index (CDAI) measures. We identified visits during follow-up where the maximum and minimum weights occurred and defined weight change and CDAI change as the differences of these measures at these visits. We defined disease activity improvement as CDAI decrease of 5 and clinically relevant weight loss as 5 kg. We performed logistic regression analyses to establish the association between improved disease activity and weight loss and baseline BMI category (≥ 25 kg/m² or <25 kg/m²). We built linear regression models to investigate the association between continuous weight loss and CDAI change among patients who were overweight/obese at baseline and who lost weight during follow-up.

Results—We analyzed data from 174 RA patients with a median follow-up of 1.9 years (IQR 1.3–2.4); 117 (67%) were overweight/obese at baseline, and 53 (31%) lost 5 kg during follow-up. Patients who were overweight/obese and lost 5 kg had three-fold increased odds of disease activity improvement compared to those who did not (OR 3.03, 95%CI 1.18–7.83). Among those

Correspondence and reprint requests: Jeffrey A. Sparks, MD, MMSc, Department of Medicine, Division of Rheumatology, Immunology and Allergy, Brigham and Women's Hospital, 60 Fenwood Road, #6016U, Boston, MA 02115, Phone: 617-525-1038, Fax: 617-713-3030, jasparks@partners.org.

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

who were overweight/obese at baseline, each kilogram weight loss was associated with CDAI improvement of 1.15 (95%CI 0.42–1.88). Our study was limited by using clinical data from a single center without fixed intervals for assessments.

Conclusion—Clinically relevant weight loss (≥ 5 kg) was associated with improved RA disease activity in the routine clinical setting. Further studies are needed for replication and to evaluate the effect of prospective weight loss interventions on RA disease activity.

Keywords

obesity; rheumatoid arthritis; weight loss; disease activity

INTRODUCTION

Obesity, defined by the World Health Organization as body mass index (BMI) $\geq 30 \text{ kg/m}^2$, is an epidemic that currently affects 34.9% of adults in the U.S. (1, 2). Obesity increases the risk of developing chronic diseases such as hypertension, diabetes mellitus, coronary heart disease, and rheumatoid arthritis (RA) (3, 4).

The relationship between obesity, weight loss, and disease control has been investigated among patients already diagnosed with chronic diseases. Health benefits of weight loss (ranging from 1.3–6.4 kg) in chronic diseases include improvements in cholesterol levels, decreased risk of cardiovascular events, and improved quality of life (5–7). Among patients with hypertension, patients that lost ≥ 5 kg had improved blood pressure (8). Weight loss of 5% was associated with improved glycemic control in patients with diabetes mellitus (9, 10). In patients with osteoarthritis, weight loss of 5% was associated with improvement in joint symptoms (11). Weight loss of 5% was also associated with low/remission disease activity in patients with psoriatic arthritis (12). However, the effect of weight loss on RA disease activity has been less studied. Most prior studies in RA investigating differences in disease activity compared patients with obesity to patients with normal BMI.

Previous studies performed among patients with RA have associated obesity with increased disease activity scores using 28 joints (DAS28) as well as worsened modified health assessment questionnaire (mHAQ). DAS28 is a validated measure of RA disease activity that combines tender and swollen joint count with laboratory measures of inflammation, either erythrocyte sedimentation rate (ESR) or C-reactive protein (CRP) (13). These serum inflammatory markers are known to be higher in obese individuals (14). Thus, it is possible that obese patients have increased DAS28 scores on the basis of obesity-related elevations of ESR or CRP independent of RA disease activity (15). Using an RA disease activity measure that does not include these serum inflammatory markers, such as the clinical disease activity index (CDAI), may therefore be preferred over DAS28 when comparing RA patients with obesity to those without obesity.

The association of obesity with worsened RA outcomes compared to those with normal BMI suggests that weight loss may improve RA disease activity. A prior study reported marked improvements in RA disease activity after bariatric surgery (16). However, these findings may not apply to other patients with RA in the routine clinical setting, since the magnitude

of weight loss after bariatric surgery is greater than would be expected for non-surgical weight loss (16). Therefore, we aimed to investigate the effect of weight loss on RA disease activity in a routine clinical setting, using the CDAI a disease activity measure that does not rely on serum inflammatory markers. We hypothesized that weight loss would be associated with improved RA disease activity.

METHODS

Study Population

We identified patients with RA in the electronic medical record at a single academic medical center, Brigham and Women's Hospital (Boston, Massachusetts), using a previously validated algorithm (17). Medical record review confirmed RA according to the 2010 ACR/EULAR classification criteria (18). All aspects of the study were approved by the Partners HealthCare Institutional Review Board.

Study Design

We performed a retrospective cohort study of patients with RA that had at least two CDAI measures obtained during routine clinical care with corresponding weight measures at those clinic visits. CDAI is a measure of RA disease activity that quantifies RA disease activity on a scale of 0–76 (higher score indicating higher disease activity), and is calculated by summing the tender joint count and swollen joint counts of 28 joints as well as physician and patient global assessment on a 0–10 scale (19).

Data Collection

Clinic visits for this study were identified as any encounter that had CDAI measured with corresponding weight measures also recorded within one week of that CDAI. Since our goal was to assess both weight and disease activity changes, all patients in this study were required to have at least two clinic visits in order to be included. For patients with more than two clinic visits with CDAI and weight measures, data were collected at each eligible clinic visit. Data on exposures, outcomes, and covariates were collected at each clinic visit using detailed medical record review.

Definitions of Weight and CDAI Changes

To calculate weight change, we selected the maximum and minimum weights (in kilograms) recorded for each subject among all visit dates that CDAI and weight were measured. We determined which of the maximum and minimum weights occurred first to order them as “initial” or subsequent.” We calculated weight change as: Weight = Subsequent weight – Initial weight. For example, if the maximum weight during the entire follow-up occurred at the second visit and the minimum weight during the entire follow-up occurred at the third clinic visit, then the weight change was calculated between these two visits. Since the maximum weight was at the initial visit, this patient would have Weight<0 and would be treated as having lost weight. Weight at any other clinic visit besides the maximum and minimum weight visits was not included in the calculations. If all weights were identical, the patient was deemed as having no weight change (Weight=0) and we considered the first and last visits to calculate the corresponding CDAI change. If there were multiple visits with

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

the same minimum or maximum weights, we chose the visits which maximized follow-up duration. We similarly defined the CDAI change as the difference between CDAI scores at the same clinic visits used to define weight change. Figure 1 further illustrates the definition of weight and CDAI change that we used in this study.

Categorization of Weight Loss

Our primary exposure was weight loss ≥ 5 kg as a binary variable among those who were overweight or obese. We chose this threshold of clinically relevant weight loss due to previously reported health benefits of weight loss above this threshold in other chronic diseases (5, 8). We used the WHO classification of BMI categories for underweight (<18.5 kg/m 2), normal weight (18.5 to <25.0 kg/m 2), overweight (25.0 to <30 kg/m 2), and obese (≥ 30.0 kg/m 2) (20). As previously described, we used the mode of the height of all measures when calculating BMI since patients may have had slightly different heights recorded at different visits (21). Since weight loss would typically only be recommended for those who were overweight or obese, we included the BMI category at baseline in our analysis.

Disease Activity Outcomes

The primary dichotomous outcome was a decrease in CDAI by ≥ 5 points. Since all measures were obtained through routine clinical care, the treating rheumatologist performed all joint counts and physician global assessments in this study. We chose this dichotomous outcome based on the minimal clinically important difference of CDAI improvement reported in prior literature (22).

Covariates

We collected covariate data at the baseline visit, including sociodemographics, lifestyles, comorbidities, RA characteristics, and medication usage. Sociodemographic covariates included age, sex, race (categorized as white or non-white), education (dichotomized as greater than high school education or high school degree or lower), smoking (dichotomized as ever or never), and BMI categories. We considered the three most prevalent comorbidities recorded: osteoarthritis, hypothyroidism, and hypertension. RA disease-specific characteristics included duration of RA, and RA serologic status (seropositive defined as positive rheumatoid factor [RF] and/or anti-cyclic citrullinated protein [anti-CCP]; seronegative as both RF and anti-CCP negative). Laboratory values included CRP. Medication use included biologic and non-biologic disease-modifying antirheumatic drugs (DMARDs), glucocorticoids, non-steroidal anti-inflammatory drugs, and opioids.

Statistical Analysis

We calculated descriptive statistics including frequencies for categorical variables, mean and standard deviation (SD) for continuous variables with normal distributions, and median, range, and interquartile range (IQR) for continuous variables with non-normal distributions. We calculated these statistics for the entire study sample and then stratified by the primary exposure of BMI category and weight loss ≥ 5 kg. We performed bivariate analyses to examine whether covariates including sex, age, RA duration, smoking status, serologic status, and steroid use were associated with the exposure and outcome. We used *t*-tests for

continuous normally distributed variables, Wilcoxon rank-sum tests for continuous non-normally distributed variables, chi-square tests for categorical variables, and Fisher's exact tests for categorical variables with small cell sizes. We evaluated these covariates as possible confounders due to their associations with RA disease activity in prior literature (23–27).

In the primary analysis investigating a threshold of weight loss, we used logistic regression to estimate the odds ratios (OR) and 95% confidence intervals (CI) for the binary outcome of improved disease activity (CDAI <5 or not) according to BMI category at baseline and weight loss of 5 kg. We also included the BMI categories of overweight/obesity and normal weight in our primary analysis, since weight loss would not be recommended to patients with normal or underweight BMI. While we did not have data on the reason for weight loss, our goal was to identify patients who may have lost weight voluntarily. Therefore, we did not analyze four patients who were normal/underweight and lost 5 kg, since weight loss would not be recommended for these patients and may have been indicative of pathologic, rather than voluntary, weight loss. The three categories in the primary exposure variable therefore consisted of: BMI $\geq 25 \text{ kg/m}^2$ and did not lose 5 kg; BMI $\geq 25 \text{ kg/m}^2$ and lost 5 kg; and BMI $<25 \text{ kg/m}^2$ and did not lose 5 kg (flow diagram illustrating sample for analysis in Figure 2).

We initially performed logistic regression analysis without adjustment. In the main analysis using a multivariable logistic regression model, we adjusted for age, sex, and baseline CDAI. Since there were only 10 outcomes in the category of BMI $\geq 25 \text{ kg/m}^2$ and lost 5 kg, we were limited in the number of covariates that we could include in a multivariable model. Therefore, we performed sensitivity analyses to examine the possible confounding effect of other variables. In these analyses, we substituted the sex variable with the following possible confounders measured at the initial visit in individual models: steroid use, DMARD use, serologic status, smoking status (ever vs. never), follow-up time, and osteoarthritis. The effect size of the BMI $\geq 25 \text{ kg/m}^2$ and lost 5 kg category was similar in all models (each of these models affected the OR by <10%), so we reported the model adjusting for age, sex, and baseline CDAI as the final multivariable model.

As a secondary analysis, we analyzed weight loss in kilograms as a continuous variable, among the subset of patients who were overweight or obese at baseline in order to investigate a dose-dependent response of weight loss. In this analysis, we used linear regression to estimate the β coefficient and 95% CI for the association between Weight and CDAI among those who were overweight/obese at baseline and lost any weight during follow-up. In the final multivariable model, we adjusted for age, sex, baseline CDAI, RA duration, smoking (ever vs. never), serologic status, steroid use at baseline, and follow-up time.

Two-sided p values <0.05 were considered statistically significant. All statistical analyses were conducted using SAS software, version 9.4.

RESULTS

We analyzed 174 RA patients that had at least two clinic visits with available CDAI and BMI measures. There were a total of 836 clinic visits, with a median of 5 visits per subject (range: 2–11) over 1.9 years (IQR 1.3–2.4) occurring between March 2012 and May 2015. Mean age at baseline was 60.4 years (standard deviation [SD] 13.2), 85% of patients were female, 85% were white, and 55% had ever smoked. The median time between the clinic visits where the minimum and maximum weights were measured and were used to calculate weight and CDAI change was 1.1 years (IQR 0.7–1.5). The most common comorbidities were osteoarthritis (68%) and hypertension (52%, Table 1). Besides CRP which was missing in 14% of patients, there were no missing data.

At baseline, the median BMI was 28.0 kg/m² (IQR 23.8–31.8) and 67% of patients were overweight or obese. The median RA duration was 9.8 years (IQR 4.1–18.6) with 78% being seropositive and 15% having deformities from RA. Seven percent of subjects were in remission, 44% had low disease activity, 29% had moderate activity, and 20% had high disease activity by CDAI at baseline. The median CDAI among those who were normal or underweight (n=53) was 11 (IQR 4–22), while the median CDAI among overweight/obese patients who did not lose 5 kg (n=93) was 10 (IQR 5–15), and 17 (IQR 7–25) for overweight/obese patients who did lose 5 kg (n=24). Within the entire study sample at baseline, 87% were on DMARDs, 59% were on a non-biologic DMARD, 61% were on a biologic DMARD, 47% were on NSAIDs, 12% were on opioids, and 32% were on glucocorticoids.

Ten out of the 24 (42%) of patients who were overweight/obese at baseline and lost above the threshold of 5 kg, had a CDAI improvement of 5 compared to 18 of the 93 (19%) patients who did not lose 5 kg. Overweight/obese patients who lost above the threshold of 5 kg had significantly increased odds of disease activity improvement compared to those who did not lose 5 kg (unadjusted OR 2.74, 95% CI 1.09–6.91). After adjustment for age, sex, and baseline CDAI, overweight/obese patients that lost 5 kg had three-fold increased odds of disease activity improvement (OR 3.03, 95% CI 1.18–7.83, Table 2) compared to those that did not lose this amount of weight. Results were similar in individual models that adjusted for steroid use, DMARD use, serologic status, smoking, follow-up time, and osteoarthritis instead of sex. Among patients who did not lose 5 kg, those with normal BMI had a trend toward improved CDAI compared to overweight/obese patients (multivariable adjusted OR 1.90, 95% CI 0.88–4.11).

Among the subset of patients who were overweight/obese at baseline and lost any weight during follow-up, there was a dose-dependent response of improved CDAI with weight loss. For every 1 kg of weight lost, CDAI improved by 1.15 points (95% CI 0.42–1.88; Figure 3), after adjusting for age, sex, RA duration, smoking, steroid use, serostatus, and follow-up time.

DISCUSSION

Among patients with RA seen in routine clinical care, we found that weight loss beyond a threshold of 5 kg was associated with a three-fold increased odds of reduced disease activity compared to overweight/obese patients that did not reach this threshold of weight loss. We further observed a dose-dependent response of weight loss and reduced disease activity amongst the subset of patients that were overweight/obese and lost any weight during follow-up. These results suggest that losing weight may improve RA disease activity. Counseling RA obese patients on the benefits of weight loss on the potential for improving disease activity may provide an additional tool for RA management to clinicians.

Prior studies have associated obesity with worsened RA outcomes compared to patients with either normal or non-obese BMI. A French study found that obese patients were 83% less likely to attain a DAS28 decrease of 1.2 after 6 months of follow-up (28). In the Better Anti-Rheumatic Farmacotherapy (BARFOT) study, obesity was associated with 50% lower odds of sustained remission by DAS28 (OR 0.51, 95% CI 0.32–0.84) and worse mHAQ scores compared to non-obese patients (13). In a meta-analysis of four previous studies, obese patients had a mean difference in DAS of 0.14 (95% CI 0.01–0.27) compared to patients with normal BMI (29). In the Epidemiologic Investigation for RA, obese patients were less likely to be in remission by DAS28 compared to RA patients with normal BMI (OR 0.58, 95% CI 0.37–0.92) (30). Similar results reporting lower odds for remission/low disease activity for obese patients compared to patients with normal weight were found in another study performed in the United Kingdom (OR 0.44, 95% CI 0.22–0.88) (31). In another meta-analysis of four previous studies, RA patients who were obese were 47% less likely to be in remission/low disease activity by DAS28 (OR 0.53, 95% CI 0.41, 0.69) (32). Most recently, in a meta-analysis of 10 prior studies, patients with obesity were 42% less likely to achieve minimal disease activity compared to patients with normal BMI (OR 0.58, 95% CI 0.40–0.85) (33). In a study performed in the Netherlands, continuous BMI was positively correlated with DAS28 ($r=0.34$, $p=0.001$) (34). In another Dutch study, obese/overweight patients had higher tender and swollen joint counts than patients with normal BMI after one year of follow-up (35). In a meta-analysis of two studies, RA patients with obesity had worse HAQ scores compared to patients with normal BMI (mean difference 0.10, 95% CI 0.01–0.19) (32).

Our results expand upon this prior literature that mostly focused on comparing BMI categories of obese and normal. Similar to this prior literature, we observed a trend towards improved disease activity for patients with normal weight compared to overweight/obese patients who did not lose weight (35–37). Additionally, we observed that patients who were overweight/obese at the baseline visit had higher CDAI measures compared to normal weight patients (29). Unlike prior studies, we aimed to investigate the effect of weight loss on disease activity, rather than static categories of BMI. We found that weight loss beyond a threshold of 5 kg was significantly associated with improved disease activity. We also included the BMI categories of overweight/obesity and normal weight in our primary analysis, since weight loss would not be recommended for patients with normal or underweight BMI. The threshold of 5 kg that we investigated has been described as an obtainable goal through diet and exercise and also improves other chronic disease outcomes

(38, 39). In the secondary analysis, we found a dose-dependent response of weight loss with improved disease activity, suggesting that weight loss beyond 5 kg may provide further improvement in RA disease activity.

While the association between obesity and worse RA outcomes is well described, few previous studies have investigated the relationship between weight change and RA disease activity. A prior study by our group investigated the effect of weight loss on RA measures after bariatric surgery (16). This retrospective cohort identified 53 subjects with RA who underwent bariatric surgery. Twelve months after weight loss surgery, patients lost a mean of 41.0 kg (SD 17.3), and 68% were in remission compared to only 26% at baseline ($p<0.01$). Another study prospectively followed a cohort of 19 patients with RA who lost a mean of 4.5 kg due to non-surgical interventions and found a significant improvement in physical function (40). In BARFOT, there was no association between weight gain during follow-up and RA disease activity, but the association of weight loss with disease activity was not reported (13).

It is possible that the findings of our study may not be generalizable to other populations since the study was performed a single site. However, characteristics of our study sample are similar to other established RA cohorts and the data were collected through routine clinical care (41, 42). Since our study was a retrospective cohort study, we used data already collected and the reason for weight loss was unknown. However, our results suggesting improved outcomes for patients with normal BMI compared to obese patients are consistent with prior literature. More patients with obesity had osteoarthritis in our study than those with normal BMI, also consistent with prior literature (43, 44). However, it is possible that improvements in pain and function from osteoarthritis after weight loss may have contributed to our findings. While we did not utilize DAS28 since this measure includes inflammatory markers that might be higher in patients with obesity, it is possible that subjective components of CDAI, such as joint tenderness, might also be biased. Therefore, even validated disease activity measures such as CDAI may not truly reflect biologic RA disease activity. We designed our study based on when CDAI was measured in the routine clinical setting, as opposed to having a standard interval for CDAI measurement. To standardize follow-up and define weight loss, we identified the maximum and minimum weights during follow-up. However, it is possible that this definition of weight loss may have influenced our results. Glucocorticoid use is associated with weight gain, leflunomide may be associated with weight loss, while other DMARDs had relatively less impact on weight change (45). However, glucocorticoid use was similar across exposure groups and was adjusted for in the analysis using weight loss as a continuous variable so is likely to influence our results. Few patients started on leflunomide during this study, so this is unlikely to explain our results. Since weight was measured in routine clinical care, it is possible that seasonal factors such as weight of clothes and mood changes might have affected our results. However, median follow-up was 1.1 years, so seasonal differences would be less likely to be important in this length of follow-up. While our goal was to capture voluntary weight loss in the routine clinical testing, we were unable to measure whether patients were actively participating in dietary or physical activity programs to lower weight. Finally, while we aimed to investigate the effect of weight loss on RA disease activity, it is possible that the converse may be responsible for the association we report.

Patients with improved RA disease activity may be more likely to lose weight through mechanisms such as improved quality of life, less pain, increased physical activity, and healthier diet. Prospective weight loss intervention studies are needed to definitively establish the causal role of weight loss and disease activity.

In conclusion, we demonstrated that weight loss above a threshold of 5 kg in overweight/obese patients was associated with a significant improvement in RA disease activity. Additionally, we found a dose-dependent response between weight loss and improved disease activity among RA patients who were overweight or obese suggesting additional benefit in disease activity for weight loss beyond 5 kg. These findings suggest that weight loss may have a role in the non-pharmacologic management to improve RA disease activity specifically for overweight or obese RA patients. Further research is needed to replicate these findings in other populations and to prospectively investigate the effect of weight loss interventions on RA disease outcomes.

Acknowledgments

Funding. This work was supported by the Rheumatology Research Foundation Scientist Development Award (Dr. Sparks) and the Student Achievement Award (Mr. Kreps) and the National Institute of Arthritis and Musculoskeletal and Skin Diseases (grant numbers K24 AR052403, K24 AR057827, P60 AR047782, L30 AR066953, R01 AR049880, P30 AR070253, P30 AR072577, and K23 AR069688). The funders had no role in study design, data collection, analysis, decision to publish, or preparation of the manuscript. The content is solely the responsibility of the authors and does not necessarily represent the official views of Harvard University, its affiliated academic health care centers, or the National Institutes of Health.

The authors would like to thank the Section of Clinical Sciences of the Division of Rheumatology, Immunology and Allergy at Brigham and Women's Hospital for helpful guidance.

References

1. Ravussin E, Swinburn BA. Pathophysiology of obesity. *Lancet*. 1992; 340(8816):404–8. [PubMed: 1353565]
2. Ogden CL, Carroll MD, Kit BK, Flegal KM. Prevalence of childhood and adult obesity in the United States, 2011–2012. *JAMA*. 2014; 311(8):806–14. [PubMed: 24570244]
3. National Task Force on the Prevention and Treatment of Obesity. Overweight, obesity, and health risk. *Archives of Internal Medicine*. 2000; 160(7):898–904. [PubMed: 10761953]
4. Qin B, Yang M, Fu H, Ma N, Wei T, Tang Q, et al. Body mass index and the risk of rheumatoid arthritis: a systematic review and dose-response meta-analysis. *Arthritis Res Ther*. 2015; 17:86. [PubMed: 25890172]
5. Felix HC, West DS. Effectiveness of weight loss interventions for obese older adults. *Am J Health Promot*. 2013; 27(3):191–9. [PubMed: 23286596]
6. Whelton PK, Appel LJ, Espeland MA, Applegate WB, Ettinger WH Jr, Kostis JB, et al. Sodium reduction and weight loss in the treatment of hypertension in older persons: a randomized controlled trial of nonpharmacologic interventions in the elderly (TONE). TONE Collaborative Research Group. *JAMA*. 1998; 279(11):839–46. [PubMed: 9515998]
7. Morey MC, Snyder DC, Sloane R, Cohen HJ, Peterson B, Hartman TJ, et al. Effects of home-based diet and exercise on functional outcomes among older, overweight long-term cancer survivors: RENEW: a randomized controlled trial. *JAMA*. 2009; 301(18):1883–91. [PubMed: 19436015]
8. Neter JE, Stam BE, Kok FJ, Grobbee DE, Geleijnse JM. Influence of weight reduction on blood pressure: a meta-analysis of randomized controlled trials. *Hypertension*. 2003; 42(5):878–84. [PubMed: 12975389]
9. Goldstein DJ. Beneficial health effects of modest weight loss. *Int J Obes Relat Metab Disord*. 1992; 16(6):397–415. [PubMed: 1322866]

10. Williamson DF, Thompson TJ, Thun M, Flanders D, Pamuk E, Byers T. Intentional weight loss and mortality among overweight individuals with diabetes. *Diabetes Care*. 2000; 23(10):1499–504. [PubMed: 11023143]
11. Christensen R, Bartels EM, Astrup A, Bliddal H. Effect of weight reduction in obese patients diagnosed with knee osteoarthritis: a systematic review and meta-analysis. *Ann Rheum Dis*. 2007; 66(4):433–9. [PubMed: 17204567]
12. Di Minno MN, Peluso R, Iervolino S, Russolillo A, Lupoli R, Scarpa R, et al. Weight loss and achievement of minimal disease activity in patients with psoriatic arthritis starting treatment with tumour necrosis factor alpha blockers. *Ann Rheum Dis*. 2014; 73(6):1157–62. [PubMed: 23771989]
13. Ajeganova S, Andersson ML, Hafstrom I, Group BS. Association of obesity with worse disease severity in rheumatoid arthritis as well as with comorbidities: a long-term followup from disease onset. *Arthritis Care Res (Hoboken)*. 2013; 65(1):78–87. [PubMed: 22514159]
14. Visser M, Bouter LM, McQuillan GM, Wener MH, Harris TB. Elevated C-reactive protein levels in overweight and obese adults. *JAMA*. 1999; 282(22):2131–5. [PubMed: 10591334]
15. Sparks CR, Moots RJ, Goodson NJ. OP0196 Obesity and Disease Activity in A Large International Rheumatoid Arthritis Cohort. *Annals of the Rheumatic Diseases*. 2014; 73(Suppl 2):136–7.
16. Sparks JA, Halperin F, Karlson JC, Karlson EW, Bermas BL. Impact of Bariatric Surgery on Patients With Rheumatoid Arthritis. *Arthritis Care Res (Hoboken)*. 2015; 67(12):1619–26. [PubMed: 26018243]
17. Liao KP, Cai T, Gainer V, Goryachev S, Zeng-treitler Q, Raychaudhuri S, et al. Electronic medical records for discovery research in rheumatoid arthritis. *Arthritis Care Res (Hoboken)*. 2010; 62(8):1120–7. [PubMed: 20235204]
18. Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, Bingham CO 3rd, et al. 2010 Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. *Arthritis Rheum*. 2010; 62(9):2569–81. [PubMed: 20872595]
19. Anderson J, Caplan L, Yazdany J, Robbins ML, Neogi T, Michaud K, et al. Rheumatoid arthritis disease activity measures: American College of Rheumatology recommendations for use in clinical practice. *Arthritis Care Res (Hoboken)*. 2012; 64(5):640–7. [PubMed: 22473918]
20. Nguyen DM, El-Serag HB. The epidemiology of obesity. *Gastroenterol Clin North Am*. 2010; 39(1):1–7. [PubMed: 20202574]
21. Baker JF, Billig E, Michaud K, Ibrahim S, Caplan L, Cannon GW, et al. Weight Loss, the Obesity Paradox, and the Risk of Death in Rheumatoid Arthritis. *Arthritis Rheumatol*. 2015; 67(7):1711–7. [PubMed: 25940140]
22. Curtis JR, Yang S, Chen L, Pope JE, Keystone EC, Haraoui B, et al. Determining the Minimally Important Difference in the Clinical Disease Activity Index for Improvement and Worsening in Early Rheumatoid Arthritis Patients. *Arthritis Care Res (Hoboken)*. 2015; 67(10):1345–53. [PubMed: 25988705]
23. Sokka T, Toloza S, Cutolo M, Kautiainen H, Makinen H, Gogus F, et al. Women, men, and rheumatoid arthritis: analyses of disease activity, disease characteristics, and treatments in the QUEST-RA study. *Arthritis Res Ther*. 2009; 11(1):R7. [PubMed: 19144159]
24. Ruban TN, Jacob B, Pope JE, Keystone EC, Bombardier C, Kuriya B. The influence of age at disease onset on disease activity and disability: results from the Ontario Best Practices Research Initiative. *Clin Rheumatol*. 2016; 35(3):759–63. [PubMed: 26245721]
25. Madsen SG, Danneskiold-Samsøe B, Stockmarr A, Bartels EM. Correlations between fatigue and disease duration, disease activity, and pain in patients with rheumatoid arthritis: a systematic review. *Scand J Rheumatol*. 2015;1–7. [PubMed: 26303149]
26. Serdaroglu M, Cakirbay H, Deger O, Cengiz S, Kul S. The association of anti-CCP antibodies with disease activity in rheumatoid arthritis. *Rheumatol Int*. 2008; 28(10):965–70. [PubMed: 18418600]
27. Haraoui B, Jovaisas A, Bensen WG, Faraawi R, Kelsall J, Dixit S, et al. Use of corticosteroids in patients with rheumatoid arthritis treated with infliximab: treatment implications based on a real-world Canadian population. *RMD Open*. 2015; 1(1):e000078. [PubMed: 26509071]

28. Ottaviani S, Gardette A, Tubach F, Roy C, Palazzo E, Gill G, et al. Body mass index and response to infliximab in rheumatoid arthritis. *Clin Exp Rheumatol*. 2015; 33(4):478–83. [PubMed: 25962513]
29. Stavropoulos-Kalinoglou A, Metsios GS, Panoulas VF, Nevill AM, Jamurtas AZ, Koutedakis Y, et al. Underweight and obese states both associate with worse disease activity and physical function in patients with established rheumatoid arthritis. *Clin Rheumatol*. 2009; 28(4):439–44. [PubMed: 19096748]
30. Sandberg ME, Bengtsson C, Kallberg H, Wesley A, Klareskog L, Alfredsson L, et al. Overweight decreases the chance of achieving good response and low disease activity in early rheumatoid arthritis. *Ann Rheum Dis*. 2014; 73(11):2029–33. [PubMed: 24818635]
31. Ellerby N, Matthey DL, Packham J, Dawes P, Hider SL. Obesity and comorbidity are independently associated with a failure to achieve remission in patients with established rheumatoid arthritis. *Ann Rheum Dis*. 2014; 73(11):e74. [PubMed: 25107560]
32. Liu Y, Hazlewood GS, Kaplan GG, Eksteen B, Barnabe C. The Impact of Obesity on Remission and Disease Activity in Rheumatoid Arthritis: A Systematic Review and Meta-Analysis. *Arthritis Care Res (Hoboken)*. 2016
33. Lupoli R, Pizzicato P, Scalera A, Ambrosino P, Amato M, Peluso R, et al. Impact of body weight on the achievement of minimal disease activity in patients with rheumatic diseases: a systematic review and meta-analysis. *Arthritis Res Ther*. 2016; 18(1):297. [PubMed: 27964760]
34. Klaasen R, Wijbrandts CA, Gerlag DM, Tak PP. Body mass index and clinical response to infliximab in rheumatoid arthritis. *Arthritis Rheum*. 2011; 63(2):359–64. [PubMed: 21279992]
35. Heimans L, van den Broek M, le Cessie S, Siegerink B, Riyazi N, Han KH, et al. Association of high body mass index with decreased treatment response to combination therapy in recent-onset rheumatoid arthritis patients. *Arthritis Care Res (Hoboken)*. 2013; 65(8):1235–42. [PubMed: 23408767]
36. Gremese E, Carletto A, Padovan M, Atzeni F, Raffeiner B, Giardina AR, et al. Obesity and reduction of the response rate to anti-tumor necrosis factor alpha in rheumatoid arthritis: an approach to a personalized medicine. *Arthritis Care Res (Hoboken)*. 2013; 65(1):94–100. [PubMed: 22730143]
37. George MD, Baker JF. The Obesity Epidemic and Consequences for Rheumatoid Arthritis Care. *Curr Rheumatol Rep*. 2016; 18(1):6. [PubMed: 26739963]
38. Bray GA. Lifestyle and pharmacological approaches to weight loss: efficacy and safety. *J Clin Endocrinol Metab*. 2008; 93(11 Suppl 1):S81–8. [PubMed: 18987274]
39. Sacks FM, Bray GA, Carey VJ, Smith SR, Ryan DH, Anton SD, et al. Comparison of weight-loss diets with different compositions of fat, protein, and carbohydrates. *N Engl J Med*. 2009; 360(9): 859–73. [PubMed: 19246357]
40. Engelhart M, Kondrup J, Hoie LH, Andersen V, Kristensen JH, Heitmann BL. Weight reduction in obese patients with rheumatoid arthritis, with preservation of body cell mass and improvement of physical fitness. *Clin Exp Rheumatol*. 1996; 14(3):289–93. [PubMed: 8809443]
41. Furst DE, Pangan AL, Harrold LR, Chang H, Reed G, Kremer JM, et al. Greater likelihood of remission in rheumatoid arthritis patients treated earlier in the disease course: results from the Consortium of Rheumatology Researchers of North America registry. *Arthritis Care Res (Hoboken)*. 2011; 63(6):856–64. [PubMed: 21337725]
42. Prince FH, Bykerk VP, Shadick NA, Lu B, Cui J, Frits M, et al. Sustained rheumatoid arthritis remission is uncommon in clinical practice. *Arthritis Res Ther*. 2012; 14(2):R68. [PubMed: 22429277]
43. Suri P, Morgenroth DC, Hunter DJ. Epidemiology of osteoarthritis and associated comorbidities. *PM R*. 2012; 4(5 Suppl):S10–9. [PubMed: 22632687]
44. Felson DT, Anderson JJ, Naimark A, Walker AM, Meenan RF. Obesity and knee osteoarthritis. The Framingham Study. *Ann Intern Med*. 1988; 109(1):18–24. [PubMed: 3377350]
45. Baker JF, Sauer BC, Cannon GW, Teng CC, Michaud K, Ibrahim S, et al. Changes in Body Mass Related to the Initiation of Disease-Modifying Therapies in Rheumatoid Arthritis. *Arthritis Rheumatol*. 2016; 68(8):1818–27. [PubMed: 26882094]

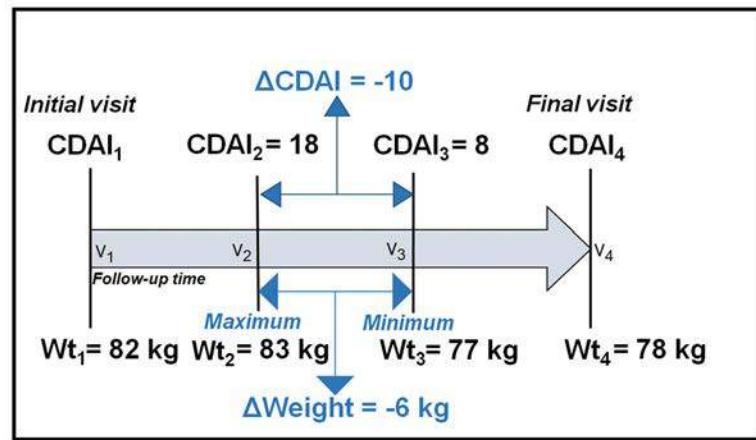
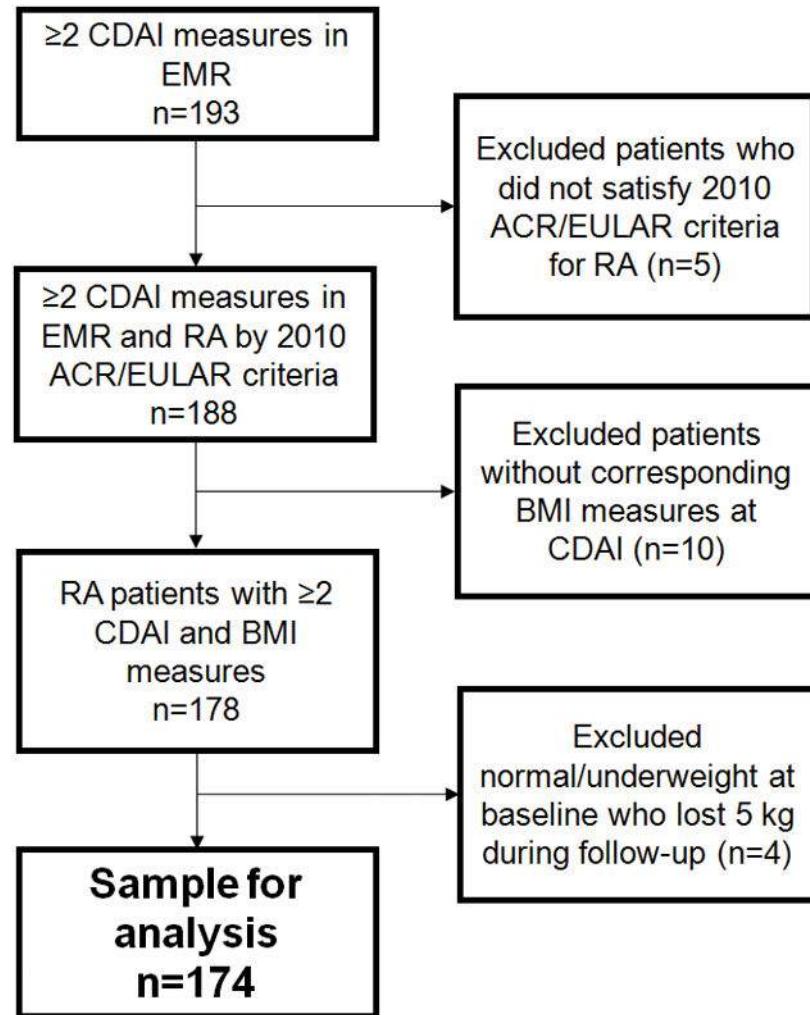


Figure 1.

Example demonstrating how weight change was defined in the study. In this patient, the maximum weight (83 kg) and minimum weight (77 kg) were selected from the four clinic visits (v₁-v₄) identified with CDAI and weight measured. The change in weight between these maximum and minimum visits was a loss of 6 kg. The change in CDAI was calculated between these same clinic visits. In this example, the CDAI change was an improvement of 10. (CDAI, clinical disease activity index; v_n, visit where a weight measure and CDAI measure were available; Wt, weight in kilograms)

**Figure 2.**

Flow diagram illustrating the identification of the final analyzed study sample (ACR, American College of Rheumatology; BMI, body mass index; CDAI, clinical disease activity index; EMR, electronic medical record; EULAR, European League Against Rheumatism; RA, rheumatoid arthritis)

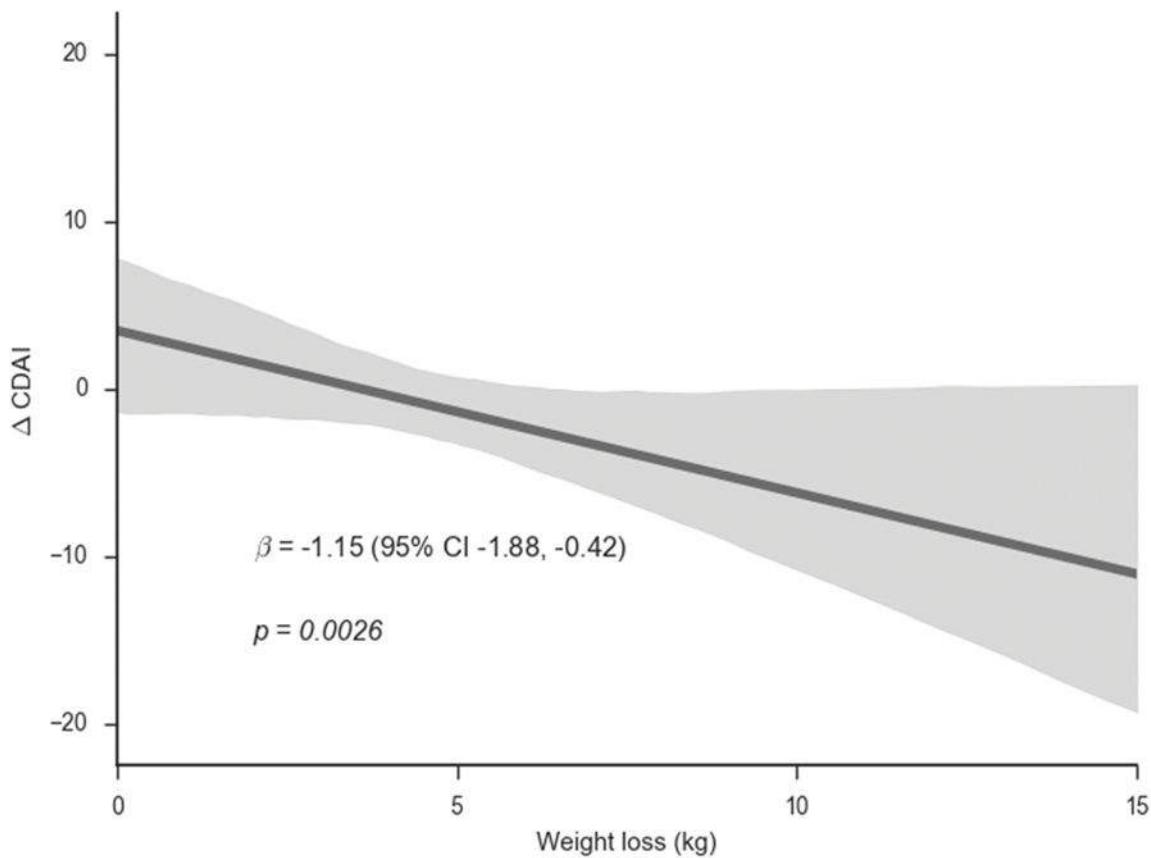


Figure 3.

Regression line and 95% confidence bounds of weight loss (continuous, in kilograms) vs. CDAI among patients with rheumatoid arthritis who were overweight or obese at baseline and lost any weight* during follow-up (n=53).

*Weight loss was defined as the difference between the maximum and minimum weight (in kg) measures at routine clinic visits and included all patients with weight <0. CDAI was calculated using measures at the corresponding clinic visits.

The linear regression model was adjusted for age, sex, baseline CDAI, RA duration, smoking (ever vs. never at baseline), serologic status (seropositive vs. seronegative), steroid use (ever vs. never at baseline), and follow-up time. (CDAI, clinical disease activity index; RA, rheumatoid arthritis).

Table 1

Baseline characteristics of patients according to BMI category and weight loss during follow-up (n=174).

Characteristic	All patients (n=174)	BMI $\geq 25 \text{ kg/m}^2$ and did not lose 5 kg (n=93)	BMI $\geq 25 \text{ kg/m}^2$ and lost 5 kg (n=24)	BMI $< 25 \text{ kg/m}^2$ and did not lose 5 kg (n=57)
Sociodemographics and Lifestyle				
Mean age, years (SD)	60.4 (13.2)	58.6 (12.9)	60.7 (12.3)	63.2 (13.7)
Female, no. (%)	147 (84.5)	73 (78.5)	19 (79.2)	55 (96.5)
White, no. (%)	148 (85.1)	79 (84.9)	20 (83.3)	49 (85.9)
>High school education, no. (%)	97 (55.8)	49 (52.7)	11 (45.8)	37 (64.9)
Ever smoker, no. (%)	96 (55.2)	52 (55.9)	16 (66.7)	28 (49.1)
Median BMI, kg/m ² (IQR)	28.0 (23.8–31.8)	29.0 (27.2–32.6)	32.4 (29.4–39.5)	23.0 (21.3–23.8)
BMI category				
Underweight, no. (%)	1 (0.6)	0 (0.0)	0 (0.0)	1 (1.8)
Normal, no. (%)	56 (32.2)	0 (0.0)	0 (0.0)	56 (98.2)
Overweight no. (%)	59 (33.9)	54 (58.1)	5 (20.8)	0 (0.0)
Obese no. (%)	58 (33.3)	39 (41.9)	19 (79.2)	0 (0.0)
Comorbidities				
Osteoarthritis, no. (%)	118 (67.8)	59 (63.4)	19 (79.2)	40 (70.2)
Hypertension, no. (%)	90 (51.7)	51 (54.8)	15 (62.5)	24 (42.1)
Hypothyroidism, no. (%)	32 (18.4)	16 (17.2)	4 (16.7)	12 (21.1)
RA characteristics				
Median RA duration, years (IQR)	9.8 (4.1–18.6)	8.3 (3.4–18.4)	10.8 (6.3–16.0)	12.2 (5.8–19.4)
Median CDAI (IQR)	10.0 (5.0–18.0)	10.0 (5.0–15.0)	16.5 (7.0–24.5)	11.0 (4.0–22.0)
RA disease activity category				
Remission (CDAI 0–2.8), no. (%)	12 (6.9)	4 (4.3)	0 (0.0)	8 (14.0)
Low (CDAI 2.9–10), no. (%)	77 (44.3)	48 (51.6)	9 (37.5)	20 (35.0)
Moderate (CDAI 10.1–22), no. (%)	51 (29.3)	28 (30.1)	8 (33.3)	15 (26.3)
High (CDAI >22), no. (%)	34 (19.5)	13 (14.0)	7 (29.2)	14 (24.6)
Components of CDAI				
Mean swollen joint count (SD)	2.9 (3.5)	2.6 (3.1)	3.0 (2.8)	3.4 (4.2)
Mean tender joint count (SD)	4.4 (5.3)	3.6 (4.5)	6.9 (7.5)	4.5 (5.0)
Mean patient global assessment (SD)	3.8 (2.4)	3.5 (2.1)	4.6 (2.7)	3.9 (2.7)
Mean physician global assessment (SD)	2.6 (1.8)	2.5 (1.9)	3.3 (1.8)	2.5 (1.7)
RF positivity, no. (%)	110 (63.2)	55 (59.1)	22 (91.7)	33 (57.9)
Anti-CCP positivity, no. (%)	82 (47.1)	48 (51.6)	12 (50.0)	22 (38.6)
Seropositive (RF or anti-CCP), no. (%)	136 (78.2)	73 (78.5)	22 (91.0)	41 (71.9)
Mean CRP, mg/L (SD) *	7.1 (10.4)	7.1 (10.1)	6.4 (5.8)	7.2 (12.6)
Medications				
Any DMARD, no. (%)	151 (86.8)	84 (90.3)	20 (80.0)	47 (82.5)
Methotrexate, no. (%)	86 (49.4)	51 (54.8)	13 (54.2)	22 (38.6)
Non-biologic DMARD, no. (%)	103 (59.2)	60 (64.5)	14 (58.3)	29 (50.9)

Characteristic	All patients (n=174)	BMI $\geq 25 \text{ kg/m}^2$ and did not lose 5 kg (n=93)	BMI $\geq 25 \text{ kg/m}^2$ and lost 5 kg (n=24)	BMI $<25 \text{ kg/m}^2$ and did not lose 5 kg (n=57)
Biologic DMARD, no. (%)	106 (60.9)	56 (60.2)	17 (70.8)	33 (57.9)
2 DMARDs, no. (%)	64 (36.8)	37 (39.8)	11 (45.8)	16 (28.1)
Glucocorticoids, no. (%)	55 (31.6)	29 (31.2)	9 (37.5)	17 (29.8)
NSAIDs, no. (%)	82 (47.1)	44 (47.3)	10 (41.7)	28 (49.1)
Opioids, no. (%)	21 (12.1)	14 (15.1)	4 (16.7)	3 (5.3)

* CRP was missing in 25 patients. All other variables had no missing values.

Anti-CCP, anti-cyclic citrullinated peptide; BMI, body mass index; CDAI, clinical disease activity index, CRP, C-reactive protein; DMARDs, disease modifying anti-rheumatic drugs; ESR, erythrocyte sedimentation rate; NSAIDs, non-steroidal anti-inflammatory drugs; RA, rheumatoid arthritis; RF, rheumatoid factor.

Table 2

Odds ratios for RA disease activity improvement (5 point improvement in CDAI) according to BMI at baseline and significant weight loss (5 kg) (n=174).

Baseline BMI and weight loss	Outcomes / Total in category (n)	% with CDAI improvement 5 points	Unadjusted OR (95% CI)	Adjusted OR (95% CI)*
BMI $\geq 25 \text{ kg/m}^2$ and did not lose 5 kg	18 / 93	19.4%	1.0 (Ref)	1.0 (Ref)
BMI $\geq 25 \text{ kg/m}^2$ and lost 5 kg	10 / 24	41.7%	2.74 (1.09–6.91)	3.03 (1.18–7.83)
BMI $<25 \text{ kg/m}^2$ and did not lose 5 kg	19 / 57	33.3%	2.12 (0.99–4.54)	1.90 (0.88–4.11)

Weight loss was defined as the difference between the maximum and minimum weights (kg) at routine clinical visits. CDAI was calculated using measures at these corresponding clinic visits and CDAI <-5 was considered a RA disease activity improvement.

Patients who had baseline BMI of $<25 \text{ kg/m}^2$ and lost 5 kg (n=4) were not analyzed. None of these patients had CDAI improvement 5 points.

* Adjusted for age, sex, and baseline CDAI.

BMI, body mass index; CDAI, Clinical Disease Activity Index; CI, confidence interval; OR, odds ratio; RA, rheumatoid arthritis.

Hepatic Transcriptome Profiles Differ Among Maturing Beef Heifers Supplemented with Inorganic, Organic, or Mixed (50 % Inorganic:50 % Organic) Forms of Dietary Selenium

James C. Matthews · Zhi Zhang · Jennifer D. Patterson · Phillip J. Bridges · Arnold J. Stromberg · J. A. Boling

Received: 18 April 2014 / Accepted: 16 June 2014 / Published online: 5 July 2014
© Springer Science+Business Media New York 2014

Abstract Selenium (Se) is an important trace mineral that, due to deficiencies in the soil in many parts of the USA, must be supplemented directly to the diet of foraging cattle. Both organic and inorganic forms of dietary Se supplements are available and commonly used, and it is known that Se form affects tissue assimilation, bioavailability, and physiological responses. However, little is known about the effects of form of dietary Se supplements on gene expression profiles, which ostensibly account for Se form-dependent physiological processes. To determine if hepatic transcriptomes of growing beef (Angus-cross) heifers (0.5 kg gain/day) was altered by form of dietary supplemental Se, none (Control), or 3 mg Se/day as inorganic Se (ISe, sodium selenite), organic (OSe, Sel-Plex®), or a blend of ISe and OSe (1.5 mg:1.5 mg, Mix) Se was fed for 168 days, and the RNA expression profiles from biopsied liver tissues was compared by microarray analysis. The relative abundance of 139 RNA transcripts was affected by Se treatment, with 86 of these with complete gene annotations. Statistical and bioinformatic analysis of the annotated RNA transcripts revealed clear differences among the four Se treatment groups in their hepatic expression profiles, including (1) solely and commonly affected transcripts; (2) Control and OSe profiles being more similar than Mix and ISe treatments; (3) distinct OSe-, Mix-, and ISe-Se treatment-induced “phenotypes” that possessed both common and unique predicted physiological capacities; and (4) expression of three microRNAs were uniquely sensitive to OSe, ISe, or Mix

treatments, including increased capacity for redox potential induced by OSe and Mix Se treatments resulting from decreased expression of MiR2300b messenger RNA. These findings indicate that the form of supplemental dietary Se consumed by cattle will affect the composition of liver transcriptomes resulting, presumably, in different physiological capacities.

Keywords Bovine · Liver · Microarray · MicroRNA · Nutrient-gene interaction · Selenium supplementation

Abbreviations

ANOVA	Analysis of variance
BOLA-	<i>Bos taurus</i> major histocompatibility complex,
DQA2	class II, DQ alpha 2
DEG	Differentially expressed gene
DDHD1	Phosphatidic acid-prefering phospholipase A1
GHRH	Growth hormone-releasing hormone
GPX1	Cellular glutathione peroxidase-1
ISe	Inorganic selenium
miRNA	microRNA
mRNA	Messenger RNA
Mix	50:50 ISe:OSe
OSe	Organic selenium
QPCT	Glutaminyl-peptide cyclotransferease
Se	Selenium
SeCys	Selenocysteine
SeMet	Selenomethionine
SEPW	Selenoprotein W

Introduction

A variety of studies have shown the negative effects of selenium (Se) deficiency in animal diets [1], including the

J. C. Matthews (✉) · Z. Zhang · J. D. Patterson · P. J. Bridges ·
J. A. Boling
Department of Animal and Food Sciences, University of Kentucky,
Lexington, KY 40546, USA
e-mail: jmatthew@uky.edu

A. J. Stromberg
Department of Statistics, University of Kentucky, Lexington,
KY 40546, USA

development of various diseases in ruminants [2] and altered metabolic and immune system capacities [3]. For beef cattle, the Se requirement of all ages is 0.10 mg Se/kg of diet daily [4]. The Se content in forages (and grains) of the southeast USA (including Kentucky) is low (<0.05 mg Se/kg) to variable (<0.10 mg/kg) [5], due to Se-poor soils. Consistently, a survey of whole blood Se concentrations of beef cows and heifers from 253 cow-calf operations in 18 states found that 42 % of southeastern USA (including Kentucky) cattle were Se deficient ($\leq 0.080 \mu\text{g/mL}$), as opposed to 18 % nationally [6]. Therefore, it is a common management practice to supplement the diets of cattle with Se in regions with Se-poor soils.

Supplementation of diets with Se is regulated by the Food and Drug Administration and cannot exceed 120 mg Se/kg or an intake of 3 mg per day when delivered in free-choice mineral supplements to cattle [7]. Diets can be supplemented with inorganic or organic forms of Se. Inorganic forms of Se typically are used (i.e., sodium selenite and sodium selenite) despite the fact that Se primarily is present in plants as organic forms. Organic forms of Se include selenomethionine (SeMet) and selenocysteine (SeCys). However, commercial animal diets are not directly supplemented with SeMet or SeCys. Instead, SeCys and SeMet are constituents of Se-enriched yeast extracts (e.g., SeMet and SeCys are major forms of Se in Sel-Plex (Alltech Biotechnologies, Inc., Nicholasville, KY, USA)) [8]. The effect of supplementing diets with inorganic versus organic Se supplements on Se bioavailability (concentrations of Se in blood and tissues) have been compared. For example, assimilation of Se after 105–106 days by whole blood, red blood cells, and liver tissue was greater for slow-maturing beef heifers fed 3 mg/day of Se as Sel-Plex than for sodium selenite [9] or from 28 to 224 days [10]. Similarly, skeletal muscle Se concentrations ($\mu\text{g/g}$) in cows supplemented with 3 mg/day as Se yeast was 1.5- to 2.5-fold higher than skeletal muscle of cows supplemented with sodium selenite [11].

To better understand the effect of supplementing beef cattle diets with different forms of Se on hepatic function, it is important to understand whether hepatic gene expression is sensitive to different forms of supplemental Se. Previously, we examined the effects of organic versus inorganic 3 mg Se/day supplementation for 105–106 days on hepatic gene expression profile of beef cattle [9] and found that organic and inorganic Se supplements commonly and differentially affected the hepatic gene expression compared to nonsupplemented animals. These Se form-induced differences corresponded with 18 and 59 % increases in hepatic content of Se by the inorganic and organic Se supplements. However, because only a single sample point (days 105/106) was used, it was not known if stabilization of hepatic Se assimilation had been achieved. Thus, conclusions about whether observed Se form induced difference in liver gene expression profiles

represented stable or transient changes could not be drawn. To address this issue and, importantly, to determine the effect of a mix of inorganic and organic forms (Mix) consisting of 1.5 mg Se/day as sodium selenite and 1.5 mg Se/day as Sel-Plex®, a second experiment using the same animal model (maturing beef heifers gaining 0.5 kg/day on a cottonseed hull-based diet) was conducted over a 224-day period using ($n=10$) none (Control) or 3 mg Se/day dietary supplement treatments as sodium selenite (ISe), Sel-Plex® (OSe), or Mix to determine the effect of Se supplement form on tissue assimilation of Se [10]. Se assimilation ($\mu\text{g/g}$) by liver tissue of ISe, OSe, and Mix treatment heifers was maximal and stable from day 56 through day 224 for all treatment groups, but the amount of Se assimilated by liver tissue was dependent on the Se treatment (OSe=Mix>ISe>nonsupplemented Control heifers).

Using the liver tissue collected from these [10] maturing beef heifers that had assimilated Se supplement form-specific amounts of Se, the objectives of the current work were (1) to compare hepatic gene transcriptome profiles when hepatic liver assimilation was stable (day 168) using microarray methods and (2) to conduct bioinformatic analysis of these profiles to predict altered hepatic physiological capacities induced by specific forms of Se supplement.

Materials and Methods

Source of Liver Tissue, Experimental Treatments, and Liver Se Concentrations

The liver tissue samples were taken by biopsy on treatment day 168 of the trial of Brennan et al. [10], for which all animal procedures were approved by the University of Kentucky Institutional Animal Care and Use Committee. Specific descriptions of the animal model, liver biopsy procedures used, and liver Se concentrations of the biopsied samples have been described [10]. However, briefly, at the end of common Se-depletion and diet-adaptation periods, 40 Angus-cross heifers were ranked (in groups of four) on the Se concentration of their biopsied liver sample. Within a rank, heifers were randomly assigned ($n=10/\text{treatment}$) to one of four dietary Se supplementation treatment groups to stratify similar basal liver Se content among treatment groups. For the control treatment (Control), no exogenous source of Se was supplied to the basal mineral-vitamin premix. For the ISe, OSe, and 1:1 combination of ISe and OSe (Mix), the basal mineral-vitamin premix was supplemented with sodium selenite (ISe, Prince Se Concentrate; Prince Agri Products, Inc., Quincy, IL, USA), Se-enriched yeast (OSe, Sel-Plex®, Alltech Inc., Nicholasville, KY, USA), or their combination, respectively. The Control, ISe, Mix, and OSe mineral-vitamin premixes contained 1.09, 36.1, 34.74, and 34.56 mg/kg of Se, respectively, and the concentration of Se in the ISe, OSe, and Mix

did not differ ($P=0.34$) [10]. These treatment premixes (85.1 g/day) were administered by manually mixing into each animal's basal diet (which contained 0.08 mg Se/kg). Feed refusals were recorded daily (and were nominal), animals were weighed biweekly, and intake of the basal diet adjusted to ensure animals maintained an average daily gain of 0.5 kg/day. The total daily Se consumption by individual heifers of the Control, ISe, Mix, and OSe treatment groups was 0.71, 3.69, 3.58, and 3.36 mg/day during days 1 to 163 and 0.82, 3.80, 3.69, and 3.76 during days 164 to 224, respectively [10]. The Se requirement of beef cattle can be met by providing 0.1 mg/kg in the complete diet [4]. Therefore, because even the Control heifers consumed 0.71 mg Se/7.73 kg diet (days 1 to 163) or 0.82 mg Se/9.09 kg diet (days 164 to 224), the complete diet of all heifers was essentially adequate in Se [4].

After 168 days of treatment, livers were biopsied and collected samples (about 1.5 to 2.5 g) placed into foil packs, snap-frozen in liquid nitrogen, and stored at -80 °C. Liver Se concentrations were analyzed by Donald C. Mahan (The Ohio State University, Columbus, OH, USA) and differences in Se-form treatments reported [10]. Specifically, mean (\pm S.D.) Se concentrations ($\mu\text{g/g}$) in liver were greater ($P \leq 0.01$) in all three Se-supplementation treatment groups than the nonsupplemented Control group (0.264 ± 0.088), whereas liver Se concentrations of Mix heifers (0.551 ± 0.231) were greater ($P=0.01$) than for ISe heifers (0.418 ± 0.070), OSe heifers (0.520 ± 0.118) tended ($P=0.08$) to be greater than ISe heifers, and Mix heifer liver Se concentrations did not differ ($P=0.20$) from OSe heifers [10].

RNA Extraction and Analysis

For each animal, total RNA was extracted from 400 mg of frozen liver tissue using TRIzol Reagent (Invitrogen Corporation, Carlsbad, CA, USA) following the manufacturer's instructions. The purity and concentration of total RNA samples was analyzed by a NanoDrop ND-1000 Spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA), which revealed that all samples were of high purity with 260/280 absorbance ratios of 2.0–2.1 and 260/230 absorbance ratios ranging from 1.5 to 1.9. The integrity of total RNA was examined by gel electrophoresis using Agilent 2100 Bioanalyzer System (Agilent Technologies, Santa Clara, CA, USA) at the University of Kentucky Microarray Core Facility. Visualization of gel images and electropherograms showed that all RNA samples had high quality with RNA integrity numbers (RIN) being greater than 8.8 and 28S/18S rRNA absorbance ratios greater than 1.8.

Microarray Analysis

The custom WT Btau 4.0 Array (version 1) (GeneChip; Affymetrix, Inc., Santa Clara, CA) was used. Microarray

analysis was conducted according to manufacturer's standard protocol at the University of Kentucky Microarray Core Facility. Briefly, 3 μg RNA for each sample was reverse-transcribed to complementary DNA (cDNA). The cDNA (double-stranded) was then reverse-transcribed to cRNA (single-stranded) which was then labeled with biotin. The biotinylated cRNA were further fragmented and used as probes to hybridize the GeneChips in the GeneChip Hybridization Oven 640, using one chip per RNA sample. Due to the loss of one animal and failure to obtain one liver sample, the experimental observations for each treatment were as follows: $n=9$ (one missed sample), 9 (one animal died), 10, and 10 for Control, ISe, Mix, and OSe treatments, respectively.

The raw expression intensity values from the 38*.cel files from the raw methylation measurements were imported into Partek Genomics Suite software (version 6.66 beta; Partek Inc., St. Louis, MO, USA). For GeneChip background correction, the algorithm of Robust Multichip Averaging adjusted with probe length and GC oligo contents was implemented [12, 13]. The background-corrected data were further converted into expression values using quantile normalization across all the GeneChips and Median Polish summarization of multiple probes for each probe set [14]. The similarity matrix was generated by Pearson (linear) correlation. The average correlation between any pair of the 38 GeneChips was 0.975, except for two GeneChips that had correlations of about 0.943 (Mix) and 0.837 (OSe). These two chips were considered outliers and removed. Pearson correlation analysis was performed again and the similarity correlation between any two of the remaining 36 GeneChips was ≤ 0.984 . The remaining 36 *.cel files ($n=9$ for each of the Se treatments) were imported into Partek Genomics Suite software and subjected to the background correction and normalization as decided above.

GeneChip transcripts were annotated using the NetAffx annotation database for Gene Expression on Bovine GeneChip Array, provided by the manufacture and last updated on September 2012. When appropriate, incompletely annotated transcripts were subjected to GenBank analysis and annotation updated. Quality control of the microarray hybridization and data presentation was performed by MA-Plot on all gene expression values and box-plot on the control probe sets on Affymetrix GeneChips (data not shown). Principal component analysis (PCA) was conducted to determine the quality of the microarray hybridization and to visualize the general data variation among the chips [14]. Differentially expressed genes (DEG) were subjected to hierarchical clustering [14] and Ingenuity Pathway Analysis (IPA, 8.7-3203; <http://www.ingenuity.com>; Ingenuity Systems, Inc., Redwood City, CA, USA).

MicroRNA Analysis

Selected microRNA (miRNA) were imported into the miRNA database (<http://www.mirbase.org/index.shtml>) to determine their mature sequences and pubmed ID. To predict their *Bos taurus* biological target, these miRNA sequences then were imported into the TargetScan (<http://www.targetscan.org>) and searched for the presence of (1) conserved branch length sites that matched the seed region of each miRNA [15], (2) sites with mismatches in the seed region that are compensated by conserved 3' pairing [16], (3) and poorly conserved sites. The resulting putative miRNA targets then were compared to the list of DEG to identify genes that were both targets of miRNA and affected by Se treatment. These genes then were subjected to the miRNA Target Filter analysis component of IPA.

Statistical Analyses

Animal was the experiment unit and one GeneChip per animal was used to evaluate relative hepatic mRNA and miRNA content in response to Se treatment. Source of variation analysis determined the sources of variation associated with treatment and error by calculating the mean *F* ratio of all 26,303 transcripts by Partek Genomic Suite (6.66 beta version) software. Selenium treatment effects on expression of all array transcripts were determined ($P<0.01$) by analysis of variance (ANOVA) using the GLM procedure of SAS (9.2 version, SAS Inst., Inc., Cary, NC, USA). When affected, the treatment least square means of DEG were separated using the pdiff option of SAS.

Results and Discussion

Experimental Model

Little is known regarding the relationship between Se concentrations in the liver and liver gene expression, especially in cattle. The effect of Se supplementation on liver gene expression using a Se-deficient versus Se-supplemented experimental model in rats [17] and chicks [18, 19] has been reported. However, because most cow-calf operations in regions of Se-poor soils supplement enough Se in mineral mixes to avoid Se deficiency, the goal of this research was to compare the relative effects of three forms of supplemental Se (ISe, Mix, and OSe) on liver gene expression profiles using a Se-adequate experimental model.

As described in the “Materials and Methods” section, after 168 days of supplementation, the concentration of Se (μg/g) in liver tissue was 0.24 (Control), 0.40 (ISe), 0.49 (Mix), and 0.49 (OSe), and differed among Se treatments (OSe=Mix>ISE>Control) [10]. Because Se concentrations of 0.25 to

0.50 μg/g in liver tissue are considered normal [20–22], even Control heifers were Se-adequate. Thus, the animal model used for both the present and a pilot trial [9] is robust with respect to both Se-adequacy and Se form-dependent Se assimilation by liver tissue. For Control heifers, liver Se content essentially was stable throughout the trial, whereas liver Se content of ISe, OSe, and Mix heifers increased until 56 to 112 days and then were stable for the remainder of the trial [10]. Thus, Se content in the day 168 liver tissue used for microarray analysis had been stable for at least 56 days.

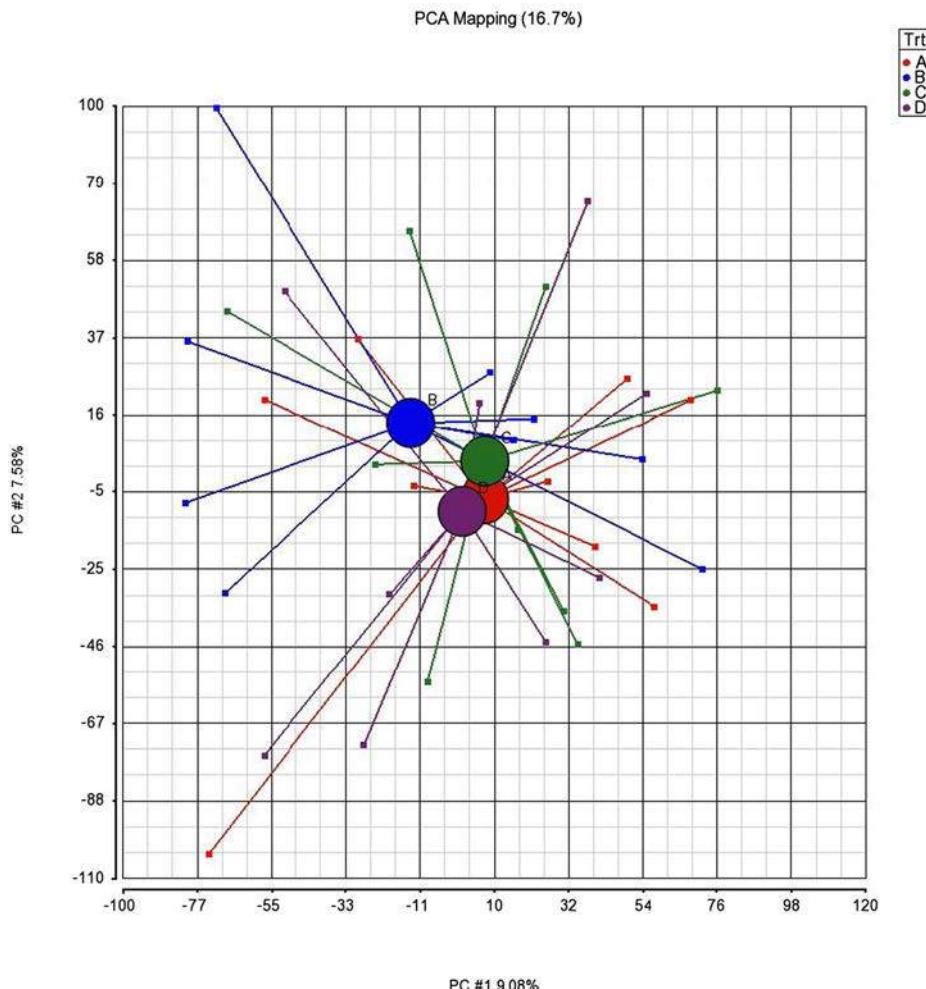
Se Treatments Altered Expression of 139 RNA Transcripts

The microarray platform used in this study was the custom WT Btau 4.0 Array (version 1, Affymetrix). This chip contains 26,303 gene transcripts (including 621 microRNA, miRNA), a complete 5' to 3' coverage of gene transcripts, and annotation based on an updated draft (4.0) of the Bovine Genome. Of note, the WT Btau 4.0 Array differs from the Bovine GeneChip Genome Array (Affymetrix) which contains 24,016 total gene transcripts, but no miRNA, and an earlier draft of the Bovine Genome. All the microarray raw data (36 *.cel files) collected with the GCOS software plus the GCRMA-processed data analyzed with the Park Genomics Suite [14] have been deposited into the Gene Expression Omnibus (National Center for Biotechnology Information, <http://www.ncbi.nlm.nih.gov/geo>) as accession number GSE44680.

Source of Variation analysis found that the mean *F* ratios for treatment and error were 0.90 and 1.00, respectively (data not shown) and principal component analysis revealed a total variance of 16.7 % (Fig. 1). The principal component (PC #1, *x* axis) contained transcripts with a higher degree of variance 9.08 %, whereas PC #2 (*y* axis) encompassed genes that had lower 7.58 % ranges of variance. In general, the GeneChips within each treatment group were clustered together and the Control, OSe, and Mix treatment groups were clustered closer than the ISe.

ANOVA identified 139 transcripts that were differentially ($P<0.01$) expressed due to form Se supplementation fed (Table 1). The difference in expression among all DEG ranged from 2.67-fold upregulated (BOLA-DQA2) to 1.65-fold downregulated (MIR2393). Hierarchical cluster analysis of the 139 DEG revealed four distinct clusters of nine GeneChips each (Fig. 2). For each cluster, all nine chips were from a single Se treatment, demonstrating distinct Se treatment effects on DEG. OSe treatment GeneChips were more closely aligned with Control GeneChips than were Mix GeneChips, and the ISe cluster was similar to Mix GeneChips and most dissimilar to Control GeneChips. These findings are consistent with those of the principal component analysis (Fig. 1).

Fig. 1 Principal component analysis of microarray transcriptome analysis of day 168 liver samples at day 168 from Control (A red), ISe (B blue), Mix (C green), and OSe (D purple) maturing Angus-cross heifers. The colored dots represent linear combinations of the relative expression data, including expression values and variances, of the 26,303 gene transcripts in each Bovine GeneChip. The center dot (centroid) for each treatment groups represents the overall treatment expression pattern



Importantly, these Se treatment-induced hepatic gene expression profiles do not match differences in hepatic Se content induced by these Se treatments (OSe=Mix>ISe>Control [10]). That is, although ISe-induced hepatic Se concentrations were more similar to Control concentrations, ISe-induced hepatic gene expression profiles were most different from Control profiles. Likewise, although the hepatic content of Se in Mix and OSe treatments were essentially identical, OSe expression profiles were more similar to Control, and Mix to ISe (Fig. 2). Thus, absolute hepatic Se assimilation was not a good indicator of gene expression. Instead, the form of Se consumed may be a better indicator. That is, the pattern of hepatic DEG was most similar between the treatments with the greatest proportion of organic Se (Control, lowest hepatic Se content) and OSe (one of the highest Se content), whereas the hepatic DEG expression profiles were similar between those animals consuming proportionately greater amounts of inorganic sources of Se (Mix and ISe). Therefore, it appears that the form of Se consumed may be more important than the amount of Se assimilated and, therefore, future trials should assess the relationship between total Se assimilation, the form of Se assimilated, and patterns of gene expression.

Common and Specific Gene Expression Profile Patterns Induced by Se Treatments

Among the 139 DEG, some were uniquely and some commonly affected by Se treatment (Fig. 3). Relative to expression by Control heifers, 32 genes were solely affected by ISe treatment, with another 22 and 4 being commonly affected by ISe and Mix or OSe treatments, respectively. In contrast, OSe supplementation solely affected the expression of 22 genes and commonly the expression of 4 and 7 genes with ISe and Mix, respectively. For the Mix treatment, 33 genes were solely affected, but another 22 and 7 were commonly expressed with ISe and OSe, respectively. In addition, 11 genes were affected commonly by all three Se supplementation treatments relative to Control, whereas the expression of another 8 genes were differentially expressed among ISe, Mix, and OSe treatments, but did not differ from Control expression levels. In total, 69 (ISe), 73 (Mix), and 44 (OSe) DEG were solely or commonly affected by Se supplementation treatments.

Although hierarchical clustering of the 139 DEG identified eight different gene expression clusters (Fig. 2), consideration

Table 1 Hepatic gene transcripts affected by form of dietary Selenium (Se) supplement to maturing beef heifers

Transcript ID	Gene symbol ^b	Gene description	Dietary Se treatments ^a				
			Control	ISe	Mix	OSe	SEM ^c
Cluster 1							
12907339	TRPC5	Transient receptor potential cation channel, subfamily C, member 5	1.00000 ^a	1.36791 ^b	1.58006 ^b	1.66591 ^b	0.12907 0.002725
12697969	LOC100139773	Hypothetical LOC100139773	1.00000 ^a	1.15171 ^{a,b}	1.36042 ^b	1.27082 ^b	0.08467 0.006423
12697805	LOC786881	Similar to hCG1812167	1.00000 ^a	-1.03848 ^a	1.05093 ^a	1.24424 ^b	0.07294 0.006400
12675620	—Intron	—	1.00000 ^{a,b}	-1.28400 ^b	1.39857 ^a	1.44240 ^a	0.18484 0.006020
12914697	LOC100336032	PREDICTED: <i>Bos taurus</i> hypothetical protein LOC100336032	1.00000 ^a	1.15312 ^{a,b}	1.13500 ^a	1.33547 ^b	0.07299 0.005211
12752300	SEPW1	Selenoprotein W, 1	1.00000 ^a	1.22275 ^b	1.21631 ^b	1.26013 ^b	0.05227 0.000289
12710910	NEK5	NIMA (never in mitosis gene a)-related kinase 5	1.00000 ^a	1.00664 ^a	1.08321 ^{a,b}	1.15446 ^b	0.04425 0.007897
12677772	—Intron	—	1.00000 ^a	1.11436 ^a	1.22302 ^a	1.60125 ^b	0.12704 0.006116
Cluster 2							
12914839	LOC100336100	PREDICTED: <i>Bos taurus</i> hypothetical protein LOC100336100	1.00000 ^a	1.15229 ^b	1.18140 ^b	1.03912 ^a	0.04393 0.001174
12881728	PCDHB14	Protocadherin beta 14	1.00000 ^a	1.28212 ^b	1.34957 ^b	1.06088 ^a	0.09183 0.006410
12864124	CYP2D14	Cytochrome P450, family 2, subfamily D, polypeptide 6	1.00000 ^{a,b}	1.15627 ^a	1.09450 ^a	-1.20399 ^b	0.08664 0.007010
12871032	RUFY3	RUN and FYVE domain containing 3	1.00000 ^a	1.02846 ^a	1.06498 ^a	-1.16205 ^b	0.05771 0.005646
12764375	LOC532807	Similar to RUN and FYVE domain containing 3	1.00000 ^a	1.09947 ^{a,b}	1.15836 ^b	-1.09996 ^a	0.06654 0.008889
Cluster 3							
12847828	MGC134040	Hypothetical protein MGC134040	1.00000 ^a	-1.03393 ^a	1.17224 ^b	-1.06679 ^a	0.06037 0.005824
12869909	CXCL2	Chemokine (C-X-C motif) ligand 2	1.00000 ^a	-1.03514 ^a	1.38450 ^b	-1.13931 ^a	0.11938 0.003537
12693465	CCNB2	Cyclin B2	1.00000 ^a	-1.06124 ^a	1.34891 ^b	-1.16518 ^a	0.11918 0.005328
12774328	GCG	Glucagon	1.00000 ^a	-1.09897 ^b	1.04313 ^a	-1.06086 ^{a,b}	0.03330 0.001132
12699461	PFN4	Profilin family, member 4	1.00000 ^a	1.20285 ^{a,b}	1.33732 ^b	-1.00647 ^a	0.09570 0.007251
12786583	IRX4	Iroquois homeobox 4	1.00000 ^a	1.03334 ^a	1.17588 ^b	-1.02677 ^a	0.04733 0.002060
12826590	LOC531024	Similar to olfactory receptor 1200	1.00000 ^a	1.06087 ^a	1.21436 ^b	1.07856 ^a	0.03869 0.000243
12686044	LOC537248	Similar to ACPL2 protein	1.00000 ^a	1.19225 ^b	1.35993 ^b	1.05331 ^{a,b}	0.07236 0.000661
12687785	RNASE12	Hypothetical protein LOC100297737	1.00000 ^a	-1.01829 ^a	1.42337 ^b	-1.05685 ^a	0.08304 0.000151
12716554	RIN2	Ras and Rab interactor 2	1.00000 ^a	-1.00644 ^a	1.09313 ^a	-1.15966 ^b	0.05574 0.001630
12782928	SNHG3-RCC1	SNHG3-RCC1 readthrough transcript	1.00000 ^a	1.07997 ^{a,b}	1.15455 ^b	1.02993 ^{a,b}	0.04166 0.009548
12827960	PLAC9	Placenta-specific 9	1.00000 ^a	1.13081 ^{a,b}	1.32674 ^b	1.10323 ^{a,b}	0.05955 0.000667
12900046	LOC510904	Hypothetical LOC510904	1.00000 ^a	1.32197 ^b	1.33335 ^b	1.23397 ^b	0.08776 0.006865
12905415	LOC539973	Similar to zinc finger, DHHC-type containing 3	1.00000 ^a	1.08064 ^a	1.32775 ^b	1.08349 ^a	0.07614 0.007144
12729265	MTMR2	Myotubularin related protein 2	1.00000 ^{a,b}	-1.02685 ^a	1.09178 ^c	1.03129 ^b	0.02711 0.001488
12706499	PSME4	Proteasome (prosome, macropain) activator subunit 4	1.00000 ^a	1.02009 ^a	1.19114 ^b	1.00366 ^a	0.05130 0.003979
12676376	—Intron	—	1.00000 ^a	1.01106 ^a	1.30861 ^b	1.32247 ^b	0.10309 0.009497
12803871	LOC782475	Similar to olfactory receptor 1368	1.00000 ^a	-1.13553 ^a	1.36319 ^b	1.09288 ^{a,b}	0.11226 0.005933
12910859	eIf4a1	Eukaryotic translation initiation factor 4A1	1.00000 ^a	1.00747 ^a	1.21265 ^b	1.10533 ^{a,b}	0.05802 0.007803
12814476	ANKRD61	Ankyrin repeat domain 61	1.00000 ^a	1.02440 ^a	1.21770 ^b	1.08275 ^a	0.05766 0.007762
12677234	—Intron	—	1.00000 ^a	1.26833 ^a	1.71743 ^b	1.12841 ^a	0.14605 0.005896
12693689	DDHD1	DDHD domain containing 1	1.00000 ^a	1.04630 ^a	1.29974 ^b	1.02109 ^a	0.06566 0.001069
12675450	—Intron	—	1.00000 ^{a,c}	-1.43052 ^b	1.21088 ^c	-1.22692 ^{a,b}	0.15143 0.008101
Cluster 4							
12677448	—Intron	—	1.00000 ^a	1.13538 ^a	-1.37222 ^b	-1.39136 ^b	0.13136 0.003895
12676664	—Intron	—	1.00000 ^a	-1.28715 ^{a,b}	-1.38955 ^{b,c}	-1.80657 ^c	0.15769 0.006690

Table 1 (continued)

Dietary Se treatments ^a							
12765783	GSG2	Germ cell associated 2 (haspin)	1.00000 ^a	-1.02219 ^a	-1.03877 ^a	-1.23345 ^b	0.06242 0.008242
12888519	LOC788287	Similar to olfactory receptor 90	1.00000 ^a	1.03204 ^a	-1.04864 ^a	-1.18675 ^b	0.05684 0.005632
12698464	QPCT	Glutaminyl-peptide cyclotransferase	1.00000 ^a	-1.16423 ^b	-1.26451 ^b	-1.29939 ^b	0.06257 0.001040
12818810	MIR2393	microRNA mir-2393	1.00000 ^a	-1.20463 ^{a,b}	-1.41205 ^{b,c}	-1.65305 ^c	0.09321 0.000119
12911831	ZNF238	Zinc finger 238 (RP58)	1.00000 ^a	1.02814 ^a	-1.10135 ^{a,b}	-1.17690 ^b	0.05856 0.008190
12686393	MFI2	Antigen p97 (melanoma associated) identified by monoclonal antigen	1.00000 ^{a,b}	1.07395 ^b	-1.06797 ^a	-1.10413 ^a	0.04999 0.009302
12676594	—Intron	—	1.00000 ^a	1.05904 ^a	1.13670 ^a	-1.26234 ^b	0.09904 0.005182
Cluster 5							
12796744	TTC21A	Tetratricopeptide repeat domain 21A	1.00000 ^a	1.21548 ^b	1.01361 ^a	1.02297 ^a	0.06029 0.007711
12912929	CAPRIN2	Caprin family member 2	1.00000 ^a	1.13153 ^b	-1.02243 ^a	-1.05430 ^a	0.04810 0.006433
12914495	LOC516089(discon)	PREDICTED: <i>Bos taurus</i> olfactory receptor 78-like	1.00000 ^a	1.25318 ^b	1.15570 ^b	1.12760 ^b	0.05707 0.004526
12879404	FGF22	Fibroblast growth factor 22	1.00000 ^a	1.29373 ^b	1.08106 ^a	1.08686 ^a	0.07032 0.006865
12873243	CNGA1	Cyclic nucleotide gated channel alpha 1	1.00000 ^a	1.19197 ^b	1.13980 ^b	1.12575 ^b	0.04591 0.004660
12857050	LALBA	Lactalbumin, alpha	1.00000 ^a	1.23641 ^b	1.20520 ^b	1.02791 ^a	0.06083 0.001969
12705807	ABO	ABO blood group (transferase A, alpha 1-3-N-acetylgalactosaminy	1.00000 ^a	1.22678 ^b	1.10102 ^{a,b}	1.02565 ^a	0.05988 0.007013
12726472	LOC100300575	Similar to olfactory receptor 154	1.00000 ^a	1.25720 ^c	1.20417 ^{b,c}	1.08372 ^{a,b}	0.05796 0.002456
12807139	LOC786846	Similar to olfactory receptor 108	1.00000 ^a	1.18358 ^b	1.14196 ^{b,c}	1.02988 ^{a,c}	0.05462 0.009475
12882374	PCDHA2	Protocadherin alpha 2	1.00000 ^a	1.16689 ^b	1.02240 ^a	-1.00409 ^a	0.04904 0.008942
12911511	ADAM7	Disintegrin and metalloproteinase domain-containing protein 7-like	1.00000 ^a	1.29932 ^b	1.16221 ^b	1.16758 ^b	0.05890 0.001454
12834487	NTM	Neurotrimin	1.00000 ^a	1.25025 ^b	1.18765 ^b	1.17953 ^b	0.06262 0.007055
12786958	ITGA2	Integrin, alpha 2 (CD49B, alpha 2 subunit of VLA-2 receptor)	1.00000 ^a	1.27304 ^b	1.22285 ^b	1.33601 ^b	0.07739 0.004259
12720027	CDH4	Cadherin 4, type 1, R-cadherin (retinal)	1.00000 ^a	1.19777 ^b	1.22550 ^b	1.10903 ^{a,b}	0.06082 0.009903
12691885	LOC614591	Similar to Olfr736 protein	1.00000 ^a	1.31777 ^b	1.07387 ^a	-1.01174 ^a	0.07211 0.001495
12747236	LOC531152	Similar to solute carrier family 5 (sodium/glucose cotran	1.00000 ^a	1.17262 ^b	1.08239 ^{a,b}	1.02624 ^a	0.04312 0.004690
12687141	LOC781919	Hypothetical LOC781919	1.00000 ^a	1.29038 ^b	1.09990 ^a	-1.02790 ^a	0.07560 0.003120
12831279	LRFN4	Leucine-rich repeat and fibronectin type III domain containin	1.00000 ^a	1.25464 ^b	1.20157 ^b	1.07978 ^b	0.06154 0.003386
12693387	LOC100337247	hCG1812119-like	1.00000 ^a	1.33301 ^b	1.22073 ^b	-1.02350 ^a	0.09509 0.005659
12885640	SHD	Src homology 2 domain containing transforming protein D	1.00000 ^a	1.17028 ^b	1.16044 ^b	1.07479 ^{a,b}	0.04725 0.006491
12843025	SLC6A17	Solute carrier family 6, member 17	1.00000 ^a	1.26507 ^b	1.23217 ^b	1.11242 ^{a,b}	0.06281 0.002701
12906697	LOC100336642	Hypothetical protein LOC100336642	1.00000 ^a	1.19347 ^b	1.23236 ^b	1.07057 ^{a,b}	0.05958 0.004127
12707153	LYG2	Lysozyme G-like 2	1.00000 ^a	1.10830 ^{a,b}	1.22910 ^b	1.05533 ^{a,b}	0.05476 0.004287
12713813	MMP9	Matrix metallopeptidase 9 (gelatinase B, 92 kDa gelatinase, 92 k	1.00000 ^a	1.24610 ^b	1.20563 ^b	1.08945 ^{a,b}	0.05874 0.002801
12842299	GJB4	Gap junction protein, beta 4, 30.3 kDa	1.00000 ^a	1.13385 ^b	1.26389 ^b	1.10493 ^{a,b}	0.05783 0.003008
12910332	LOC618023	Similar to mCG1035526	1.00000 ^a	1.13489 ^b	1.16805 ^b	1.00666 ^a	0.05158 0.006880
12829684	CEND1	Cell cycle exit and neuronal differentiation 1	1.00000 ^a	1.20278 ^b	1.10631 ^{a,b}	-1.04940 ^a	0.05132 0.000338
12911453	LOC617745	PREDICTED: <i>Bos taurus</i> histone cluster 1, H2bd-like	1.00000 ^a	1.33297 ^b	1.26593 ^b	1.02369 ^a	0.08867 0.004406
12677810	—Intron	—	1.00000 ^a	1.80027 ^b	1.44397 ^{a,b}	1.01137 ^a	0.18686 0.007738
Cluster 6							
12677714	—Intron	—	1.00000 ^a	1.25973 ^b	1.04358 ^a	1.23621 ^b	0.07479 0.006589
12802414	LOC523769	Similar to olfactory receptor 11A1 (olfactory receptor OR	1.00000 ^a	1.33145 ^b	1.17429 ^b	1.17281 ^b	0.07550 0.008159
12741683	LOC100295022	Similar to pre-B lymphocyte gene 2	1.00000 ^{a,c}	1.12993 ^{a,b}	-1.06419 ^c	1.17265 ^b	0.06415 0.005043
12740074	LOC530749	Hypothetical LOC530749	1.00000 ^a	1.18507 ^b	-1.01920 ^a	1.08987 ^{a,b}	0.05729 0.008473

Table 1 (continued)

Dietary Se treatments ^a							
12804717	GCM1	Glial cells missing homolog 1 (<i>Drosophila</i>)	1.00000 ^a	1.20950 ^b	1.04205 ^a	1.05058 ^a	0.05491 0.009211
12686542	C1H21ORF7	Chromosome 21 open reading frame 7 ortholog	1.00000 ^a	1.04049 ^a	-1.16979 ^b	-1.03255 ^a	0.05544 0.006607
12906283	MIR222	microRNA mir-222	1.00000 ^a	1.22316 ^b	-1.00025 ^a	-1.09591 ^a	0.06853 0.001913
12824553	ZNF703	Zinc finger protein 703	1.00000 ^a	1.14629 ^b	-1.04146 ^a	1.03012 ^a	0.04940 0.007012
12730430	LOC788649	Similar to odorant receptor HOR3beta3	1.00000 ^a	1.22925 ^b	-1.00493 ^a	1.07249 ^a	0.06580 0.008928
12911627	LOC790172	PREDICTED: <i>Bos taurus</i> membrane glycoprotein gp140-like	1.00000 ^a	1.25359 ^b	1.10068 ^{a,c}	1.15272 ^{b,c}	0.05657 0.006471
12713247	LOC100196900	Hypothetical LOC100196900	1.00000 ^a	1.09330 ^a	-1.00273 ^a	-1.18405 ^b	0.07243 0.008923
12866241	TULP3	Tubby-like protein 3	1.00000 ^a	1.06038 ^a	-1.11448 ^b	-1.00561 ^a	0.04430 0.007012
12876076	LOC782977	Similar to pol protein	1.00000 ^a	1.16389 ^b	-1.10204 ^a	1.06022 ^{a,b}	0.06658 0.008782
12913095	PAKD1	—	1.00000 ^a	1.74958 ^b	1.04876 ^a	1.04799 ^a	0.15558 0.003061
12902043	LOC100298746	Similar to protein phosphatase 2, regulatory subunit	1.00000 ^a	1.17319 ^b	1.00527 ^a	1.02182 ^a	0.04358 0.001923
12882371	OR7E24	Similar to olfactory receptor 18 (olfactory receptor 145-	1.00000 ^a	1.42674 ^b	1.01118 ^a	1.06376 ^a	0.08860 0.001017
12854700	FBXL13	F-box and leucine-rich repeat protein 13	1.00000 ^a	1.16934 ^b	-1.03845 ^a	1.03846 ^a	0.05034 0.003826
12717977	GHRH	Growth hormone releasing hormone	1.00000 ^a	1.22927 ^b	1.06399 ^a	-1.00427 ^a	0.06520 0.009874
12712598	PDX1	Pancreatic and duodenal homeobox 1	1.00000 ^a	1.16694 ^b	-1.02570 ^a	1.02967 ^a	0.05157 0.006236
12682969	LOC100296558	Hypothetical protein LOC100296558	1.00000 ^a	1.41912 ^b	1.15929 ^a	1.08157 ^a	0.07247 0.000266
12674910	—Intron	—	1.00000 ^a	1.35052 ^b	-1.21602 ^a	1.12980 ^a	0.11494 0.001786
12729575	HBE1	Hemoglobin, epsilon 1	1.00000 ^a	1.07921 ^{a,b}	-1.14303 ^b	-1.10060 ^{a,b}	0.06238 0.008809
Cluster 7							
12856863	LDHB	Lactate dehydrogenase B	1.00000 ^a	-1.02584 ^a	-1.13592 ^a	1.21088 ^b	0.08583 0.007247
12707574	VAMP5	Vesicle-associated membrane protein 5 (myobrevin)	1.00000 ^a	1.09221 ^{a,b}	1.01750 ^{a,b}	1.18502 ^b	0.04999 0.008498
12906833	LOC100337148	Nuclear RNA export factor 3-like	1.00000 ^a	1.20167 ^b	1.03651 ^{a,b}	1.34291 ^b	0.08945 0.008903
12675356	—Intron	—	1.00000 ^a	1.10150 ^a	-1.43145 ^b	1.14442 ^a	0.10746 0.000423
12844091	TNFAIP8L2	Tumor necrosis factor, alpha-induced protein 8-like 2	1.00000 ^a	1.02130 ^a	-1.10577 ^b	1.07832 ^a	0.04659 0.007207
12770586	CDRT4	CMT1A=duplicated region transcript 4	1.00000 ^a	-1.04470 ^{a,b}	-1.15988 ^b	1.06567 ^{a,b}	0.06043 0.009912
12676756	—Intron	—	1.00000 ^a	1.13462 ^{a,b}	-1.12408 ^a	1.22693 ^b	0.09456 0.009437
12885754	LRRTM2	Leucine-rich repeat transmembrane neuronal 2	1.00000 ^a	1.17937 ^b	-1.03318 ^a	1.12832 ^b	0.05972 0.005353
12675362	—Intron	—	1.00000 ^a	-1.15313 ^{a,b}	-1.42605 ^c	-1.24480 ^{b,c}	0.09101 0.004356
12678012	—Intron	—	1.00000 ^a	-1.00494 ^a	-1.26902 ^b	-1.10875 ^a	0.06429 0.002107
12857735	PLCZ1	Phospholipase C, zeta 1	1.00000 ^a	-1.06351 ^{a,b}	-1.19444 ^c	-1.13876 ^{b,c}	0.04580 0.002755
12910700	MCART6	Similar to mitochondrial carrier triple repeat 6	1.00000 ^a	1.05026 ^a	-1.17399 ^b	1.01611 ^a	0.06033 0.006992
12871908	FRAS1	Similar to rCG64566	1.00000 ^a	1.03201 ^a	-1.17755 ^b	-1.00326 ^a	0.05950 0.009642
12895450	SLC35D2	Solute carrier family 35, member D2	1.00000 ^a	-1.09449 ^b	-1.13431 ^b	1.00502 ^a	0.04307 0.008158
12732657	OR2AT4	Similar to olfactory receptor Olr35	1.00000 ^a	-1.03373 ^{a,b}	-1.11814 ^b	1.06801 ^a	0.04643 0.007303
12759892	LOC785933	Similar to Orphan sodium- and chloride-dependent neurotra	1.00000 ^a	1.01002 ^{a,b}	-1.14507 ^a	1.15762 ^b	0.06967 0.004500
12914365	LOC100462699	—	1.00000 ^{a,b}	1.23788 ^b	-1.26864 ^a	1.13590 ^b	0.11736 0.004271
Cluster 8							
12915559	BOLA-DQA2	<i>Bos taurus</i> major histocompatibility complex, class II	1.00000 ^a	-1.30149 ^a	-1.21918 ^a	2.66633 ^b	0.37359 0.008451
12870473	NPY1R	Neuropeptide Y receptor Y1	1.00000 ^a	-1.31851 ^b	-1.11101 ^a	1.06292 ^a	0.07942 0.001029
12851703	TMEM168	Transmembrane protein 168	1.00000 ^{a,b}	-1.04902 ^b	1.07443 ^a	1.07428 ^a	0.03587 0.005384
12677980	—Intron	—	1.00000 ^a	-1.26612 ^b	-1.11131 ^{a,b}	1.20419 ^a	0.10765 0.004477
12805856	LOC521580	Similar to histone cluster 1, H2bd	1.00000 ^a	1.03429 ^a	1.04456 ^a	1.65795 ^b	0.15039 0.004848

Table 1 (continued)

Dietary Se treatments ^a							
12913565	PLOD1	Procollagen-lysine, 2 oxoglutarate 5-dioxygenase1	1.00000 ^{a,b}	-1.16906 ^a	-1.07681 ^a	1.22793 ^b	0.10040 0.008681
12899187	LOC100297846	Similar to trace amine associated receptor 8	1.00000 ^a	-1.03889 ^a	1.06342 ^a	1.20792 ^b	0.05900 0.002913
12730199	LOC782645	Similar to olfactory receptor Olr245	1.00000 ^a	1.02972 ^a	1.12569 ^{a,b}	1.25095 ^b	0.05992 0.003096
12810320	ZNF236	Zinc finger protein 236	1.00000 ^{a,b}	-1.06458 ^a	-1.03225 ^a	1.05678 ^b	0.03261 0.007426
12878036	THG1L	tRNA-histidine guanylyltransferase 1-like (<i>S. cerevisiae</i>)	1.00000 ^a	-1.14203 ^b	-1.07103 ^{a,b}	-1.00782 ^a	0.03735 0.003804
12708470	LOC782479	Similar to ribosomal protein L26	1.00000 ^a	-1.21664 ^b	-1.06766 ^{a,b}	1.11736 ^a	0.08608 0.009312
12870694	SRP72	Signal recognition particle 72 kDa	1.00000 ^{a,b}	-1.02321 ^a	-1.08373 ^a	1.08183 ^b	0.04071 0.006627
12857895	PUS7L	Pseudouridylate synthase 7 homolog (<i>S. cerevisiae</i>)	1.00000 ^a	-1.10623 ^a	-1.09610 ^a	1.13670 ^b	0.05312 0.000423
12743540	GOLGA3	Golgi autoantigen, golgin subfamily a, 3	1.00000 ^a	-1.06641 ^{a,b}	-1.08873 ^b	1.08322 ^a	0.03868 0.000934
12842784	FCGR2	Fc fragment of IgG, low affinity IIb, receptor (CD32)	1.00000 ^a	-1.13416 ^{a,b}	-1.27149 ^b	1.13919 ^a	0.08708 0.001204
12747334	THOC5	THO complex 5	1.00000 ^a	-1.08233 ^b	-1.03511 ^{a,b}	1.03841 ^a	0.03025 0.004399
12809240	ADNP2	ADNP homeobox 2	1.00000 ^a	-1.10483 ^b	-1.01514 ^a	1.06486 ^a	0.03198 0.000331
12844243	SNAPIN	SNAP-associated protein	1.00000 ^a	-1.14259 ^b	-1.19432 ^b	-1.07056 ^{a,b}	0.05038 0.007386
12828917	ZNF22	Zinc finger protein 22 (KOX 15)	1.00000 ^a	-1.14634 ^b	-1.18867 ^b	-1.07987 ^{a,b}	0.04388 0.002467
12690686	ACIN1	Apoptotic chromatin condensation inducer 1	1.00000 ^a	-1.07579 ^b	-1.11190 ^b	-1.01875 ^{a,b}	0.02836 0.002180
12913489	ACLA_046210	DnaJ domain protein	1.00000 ^a	-1.18778 ^{a,b}	-1.40734 ^b	1.12130 ^a	0.11243 0.002185
12840769	LEPRE1	Leucine proline-enriched proteoglycan (leprecan) 1	1.00000 ^a	-1.08090 ^{a,b}	-1.11548 ^b	1.05989 ^a	0.04569 0.003257
12677698	—Intron	—	1.00000 ^a	-1.12940 ^a	-1.06740 ^a	1.15727 ^b	0.06441 0.001365
12912371	GCLC	Glutamate-cysteine ligase, catalytic subunit	1.00000 ^a	-1.24385 ^b	-1.19484 ^b	-1.23431 ^b	0.06410 0.005740
12913085	CEP350	Centromere-associated protein 350	1.00000 ^a	-1.56052 ^b	-1.38906 ^b	-1.08920 ^a	0.10649 0.000839
12708824	MIR2300B	microRNA mir-2300b	1.00000 ^a	-1.24583 ^b	-1.32703 ^b	-1.03539 ^a	0.07937 0.001808

Se supplement treatments that contained no Se (Control) or 3 mg Se/day in the form of sodium selenite (ISe), 1.5 mg Se/day of each sodium selenite and SelPlex® mix (Mix), or 3 mg Se/day of SelPlex® (OSe) were top-dressed onto enough of a common cottonseed hull/soybean hull/cracked corn-based diet (0.08 mg Se per day) to support 0.5 kg/day growth in maturing Angus-cross heifers for 168 days. The abundance of gene transcripts are reported relative to the mean expression of the Control group and are expressed as the fold change of nontransformed data

P values were obtained from one-way ANOVA F test

^a Means with different superscripted letters differ ($P < 0.05$)

^b A missing symbol indicates a lack of confirmed bovine DNA or RNA sequence annotation

^c The presented SEM values were pooled (averaged) from that of Control ($n=9$), ISe ($n=9$), Mix ($n=9$), and OSe ($n=9$) treatment groups

of the Se treatment effects on only the 86 fully annotated DEG (Table 1) reveal four prominent, overarching patterns of Se treatment-specific effects on gene expression. First, from clusters 1 (three genes) and 5 (15 genes), any Se form supplement (ISe, Mix, OSe) increased gene expression relative to Control. Second, in cluster 3 (15 genes), it is apparent that the Mix treatment upregulates gene expression, whereas OSe or ISe downregulates gene expression. Third, from cluster 6 (11 genes), OSe did not affect expression whereas ISe treatment upregulated gene expression and Mix treatment downregulated expression. Fourth, from clusters 7 (10 genes) and 8 (20 genes), either Mix or ISe treatments downregulated gene expression whereas OSe upregulated expression. In addition, two other Se form-induced patterns (clusters 2 and 4) are

represented in the gene clusters, but the relative few genes involved with each cluster reduce confidence in their validity. Specifically, from cluster 2 (three genes), it would appear that whereas OSe upregulates gene expression, ISe, or Mix decreased expression. In contrast, from cluster 4 (five genes), whereas one gene is downregulated by all three Se treatments, three others are downregulated by OSe or Mix, but another is upregulated by only ISe treatment.

Potential Physiological Consequences of DEG

Hierarchical clustering analysis identified eight distinct patterns/clusters of DEG that were sensitive to supplemental Se treatment (Fig. 2, Table 1). Bioinformatic analysis was

Fig. 2 Hierarchical cluster analysis of 139 differentially expressed ($P < 0.01$) gene transcripts by liver tissue of maturing Angus-cross heifers after 168 days of Se supplementation. Heifers consumed no supplemental Se (Trt A, red, Control) or 3 mg Se/day in the form of SelPlex® (Trt D, purple, OSe), 1.5 mg Se/day of each sodium selenite and SelPlex® mix (Trt C, green, Mix), or 3 mg Se/day as sodium selenite (Trt B, blue, ISe). Cluster analysis among chips is shown by the top color bar. The expression level for each gene transcript was standardized to mean of 0 and scaled to a standard deviation of 1, which is the default setting of the Partek Genomics Suite software. As indicated by the bottom color bar, gray in the middle represents the mean value (0), red represents gene expression levels above the mean expression, whereas blue denotes expression below the mean. The intensity of the color reflects the relative intensity level of transcript expression

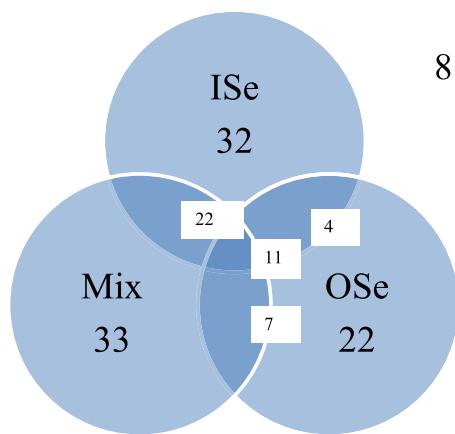
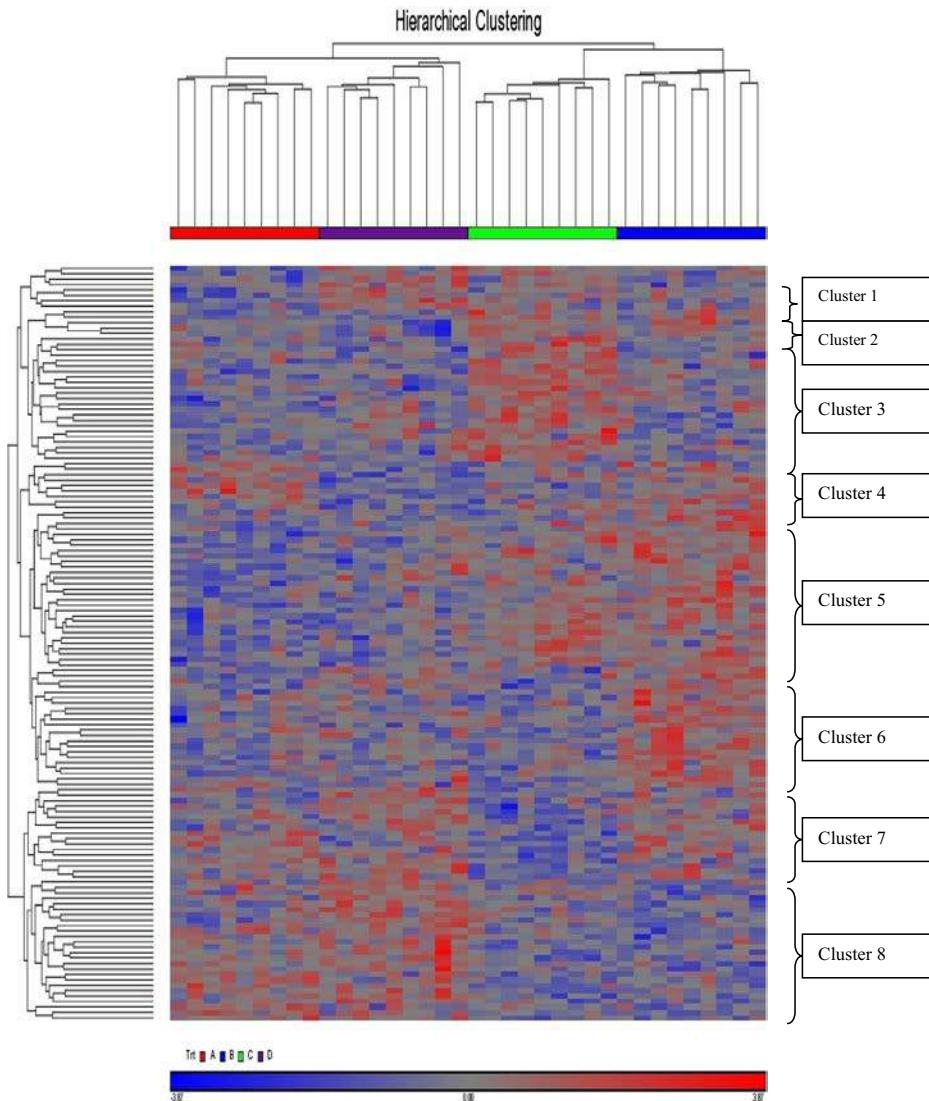


Fig. 3 Zen diagram depiction of the relationship of 139 differentially expressed gene transcripts by liver tissue of maturing Angus-cross heifers after 168 days of no supplemental Se (Control) or 3 mg Se/day in the form of sodium selenite (ISe), 1.5 mg Se/day of each sodium selenite and SelPlex® mix (Mix), or 3 mg Se/day of SelPlex® (OSe). Note that expression of eight transcripts differed among ISe, Mix, or OSe treatments, but not versus Control heifers

conducted on annotated DEG within each cluster to identify the potential physiological consequence of specific Se treatment effects within clustered DEG and the predominant (the highest ratio of DEG to all molecules) gene networks reported.

Two gene networks were identified in cluster 1, organismal disease (SEPW1) and molecular and mineral transport (TRPC5). Functional analysis identified eight transcripts involved in cation transport and antioxidative activity. SEPW1 (selenoprotein W-1) contains SeCys (SeCys¹³), encoded by the UGA codon. SEPW1 belongs to the selenoprotein W subfamily, of the selenoprotein WTH family, and functions as an antioxidant enzyme. Specifically, SEPW1 binds to glutathione which targets reactive oxygen species such as hydrogen peroxide, superoxide anion radical, and hydroxyl radical [23]. The finding that SEPW1 mRNA content was increased by all three Se-containing mineral mix treatments (Table 1) is consistent with the well-known sensitivity of SEPW1

expression to supplemental Se, and demonstrates that the experimental model was capable of responding to dietary Se supplementation. Metabolically, that ISe, OSe, and Mix all increased SEPW1 mRNA suggests that they had a greater antioxidant capacity.

As observed for SEPW1, TRPC5 mRNA expression was equally stimulated by all forms of Se supplement. Homologues of transient receptor potential (TRP) genes encode a variety of cation channels, most of which conduct Ca^{2+} across the plasma membrane [24]. Thus, TRP channels play pivotal roles in sensing and adapting to a wide variety of environmental changes [25] and are thought to have a relatively greater importance in “active” adaptive (rather than “passive”) responses to physiological challenges [26]. In this regard, TRPC5 mRNA is expressed by adult and fetal hepatocytes [27] and TRPC5 channels are known to be stimulated by binding by extracellular thioredoxin [28] and hydrogen peroxide [26]. Consequently, TRPC5-mediated Ca^{2+} influx in response to these oxidative stressors results in modulation of nutrient metabolism, including that of glucose, fatty acids, and amino acids [29]. Thus, if TRPC5 mRNA levels represent functional capacity, then Se supplementation induced an increased capacity to respond to oxidative stress in Se-supplemented heifers versus the Se-inadequate, but nonsupplemented Control heifers.

Cluster 2 DEG (CYP2D14, PCDHB14, and RUFY3) are involved with the cell-to-cell signaling pathway. PCDHB14 expression was increased by ISe and Mix treatments, respectively. In contrast, CYP2D14 and RUFY3 expression were decreased by OSe. PCDHB14 is a cadherin protein that belongs to the protocadherin β family and likely is involved in regulation of neuronal development in peripheral tissue [30]. Concomitantly, RUFY3 (Singar 1) is an early endosome-localized protein that associates with phosphatidylinositol 3-phosphate in membranes of early endosomes, and is involved in Ras-like GTPase signaling pathways through binding to Rab5 and Rab33A [31, 32]. In neurons, RUFY3 suppresses excess axons formation, thus ensuring neuronal polarity [32]. Thus, the increased expression of PCDHB14 mRNA by Mix and OSe Se treatments, and downregulation of RUFY3 by OSe, suggests an altered capacity of neuronal regulation of liver tissue.

With regard to DEG function, CYP2D14 is the bovine ortholog of CYP2D6 [33], a well-studied member of the cytochrome P450 family with xenobiotic catabolic activity. In humans at least, CYP2D6 is most highly expressed in the liver, but lowly expressed compared to other liver cytochrome P450 proteins [34]. However, CPY2D6, which demonstrates promiscuous substrate activity, is thought to metabolize 15 to 25 % of all clinical drugs [34, 35]. Moreover, because of the affinity that CYP2D6 displays for plant alkaloids, the high level of polymorphism in CYP2D6 is thought to convey an evolutionary selection advantage by facilitating a greater

variety of food sources, due to detoxify a broad spectrum of xenobiotics [35]. Although CPY2D6 is reported to be insensitive to cellular environmental or hormonal regulation [34, 35], our finding that OSe decreased CYP2D14/CYP2D6 mRNA content suggests CYP2D14 is sensitive to form of Se supplementation and that OSe heifers may possess a decreased capacity to metabolize a variety of xenobiotic compounds. Whereas this may be beneficial when undergoing certain pharmaceutical therapies, it may be a disadvantage when consuming feedstuffs with high levels of ergot alkaloids.

Cluster 3 DEG are involved in tissue development and protein synthesis. There was a marked difference in Se treatment effects on cluster 3 DEG. Specifically, Mix treatment upregulated 20 of 23 transcripts and 13 of 15 fully annotated genes in cluster 3. In contrast, ISe treatment downregulated expression of glucagon (GCG) and OSe downregulated RIN2 mRNA expression. This latter finding suggests that the capacity for Rab5-initiated signaling cascades (e.g., endocytosis of E-cadherin) is reduced in livers of OSe heifers. That is, Rab5 is a key regulator of transport from the plasma membrane to the early endosomes and RIN2 functions in cell-cell adhesion sites to bind both Rab5 and Ras, thereby enabling Rab5-mediated endocytic events [36]. However, it remains to be determined whether the OSe-induced reduction of RIN2 expression would be great enough to disrupt Rab5-mediated control of endosome organization [37].

The Mix Se treatment increased expression of phosphatidic acid-prefering phospholipase A1 (DDHD1). DDHD1 cleaves phosphatidylinositol to generate arachidonic acid-containing lysophosphatidylinositol which, after export by the ABCC1 transporter, binds to G protein-coupled 55 (GPR55) in a paracrine or autocrine manner [38]. DDHD1 [39] and GPR55 are expressed in liver tissue [40]. As summarized by others [41], DDHD1-mediated activation of GPR55 binding by lysophosphatidylinositol results in direct activation (phosphorylation) of p38 MAPK and stimulation of Ca^{2+} from endoplasmic reticulum. This increase of cytosolic Ca^{2+} results in activation of transcription factors CREB and NF- $\kappa\beta$ by ERK1/2 and translocation of NFAT into the nucleus. Because activation of the ERK1/2 and NFAT pathways results in enhanced fatty acid oxidation and uptake, and glucose uptake, at least in skeletal muscle tissue [41], the increased expression of DDHD1 by Mix Se treatment likely represents an enhanced flexibility of energy metabolism.

The translation of mRNA into protein is mediated by eukaryotic translation initiation factors (eIF). eIF4a1 is a eukaryotic transcription factor involved in the regulation of protein synthesis. More specifically, eIF4a1 has ATP-dependent helicase activity that unwinds mRNA secondary structures in the 5' untranslated region, thus facilitating 40S ribosomal subunit identification Met-tRNA. Moreover, increased eIF4a content is thought to enhance translation of

mRNA with complex secondary structures at both 5' and non-5' sites [42]. Whether the greater mRNA content of eIf4a1 in Mix heifers suggests a greater capacity for translation initiation of all proteins or proteins with mRNA that have a relatively high proportion of secondary structure remains to be determined.

Cluster 4 DEG were downregulated by Se treatments, except for MF12, which was upregulated by ISe treatment. Of the cluster 4 DEG, GSG2 was down-regulated by OSe treatment and is involved in the chromatin modification (especially that of histone phosphorylation), and helps to regulate chromosome segregation during mitosis [43]. Zinc Finger 238 (ZNF238) is a member of the POZ-zinc finger family and functions as a DNA binding protein that represses transcription [44] and also was downregulated by OSe consumption. These proteins are important regulators of DNA damage responses, cell cycle progression, and many developmental events [45]. Of the cluster 4 DEG, MiR2393 was highly affected, being downregulated by Mix (1.41-fold) and OSe (1.65-fold), and quantitatively downregulated by ISe. As with ZNF238, MiR2393 acts to repress expression, although by binding mRNA rather than DNA (see below).

Similar to MiR2392, glutaminyl-peptide cyclotransferease/glutaminyl cyclase (QPCT) mRNA expression was downregulated by all three Se treatments. QPCT is involved in the synthesis of pyroglutamyl peptides by posttranslational cleavage of nascent gene products to their respective N-terminal pyroglutamyl bioactive peptides, including thyroid-releasing hormone, gonadotrophin-releasing hormone, and corticotrophin-releasing hormone [46]. In liver tissue, QPCT activity promotes maturation and longevity of chemokines, resulting in enhanced inflammatory responses [47]. Thus, an important finding from this study is that QPTC-enhanced inflammation may be sensitive to Se supplementation.

The predominant network of cluster 5 was cell movement development. The common effect of Se supplemental treatments on the 18 cluster 5 DEG was upregulation. Four genes were commonly upregulated by all three Se supplemental treatments, eight by only ISe and Mix, and six by ISe only, and one by Mix only. Of the cluster 5 DEG, ITGA2 encodes the alpha-chain of integrins found in T cells and fibroblasts, and is involved with cell adhesion and cell-surface mediated signaling. Importantly, the integrin complex can mediate the activation of hepatic stellate cells [48]. ITGA2 upregulation indicates greater capacity for healing and IL-3 mediated immune response. ITGA2 also is involved in activation of monocytes to trigger monocyte migration to inflammatory sites to regulate IL-13-mediated monocyte activation [49]. Accordingly, the upregulation of ITGA2 in ISe, OSe, and Mix heifers indicates an increased activation of macrophages and better tissue wound healing capacity in response to injury.

SLC6A17 (NTT4/XT1) is a glutamatergic and GABAergic neuron-specific transporter localized to the plasma membrane

and responsible for Na^+ -dependent absorption of (primarily) glutamine and proline into the cytosol [50]. SLC6A17 mRNA content was equally upregulated by ISe and Mix treatments. If the bovine expression pattern for SLC6A17 is the same as other species, then this finding indicates that glutamatergic neurons which innervate the liver have an increased neural propagation function, due to the increased capacity to absorb glutamine and proline, both precursors of glutamate.

The predominant network found for cluster 6 genes was regulation of hormone metabolism. ISe treatment increased gene expression for 8 of 11 annotated cluster 6 DEG, including GHRH, Mir-222, and PDX1. In contrast, Mix treatment decreased expression of three genes (C1H21ORF7, TULP3, and HBE1), and OSe had no effect. GHRH acts directly on specific extrapituitary peripheral tissues to induce cell proliferation by binding to the GHRH receptor, SV-1 [51, 52]. Both SV-1 mRNA and protein is expressed by mouse liver tissue [52] and GHRH expression was found in mouse tissues that express SV-1, except for the liver [53]. Extrahypothalamic GHRH expression has been reported in the gonads [54, 55], pancreas [56], placenta [55], and, by this report, the liver. Our finding that more GHRH mRNA expression was increased in ISe liver tissue suggests an increased hepatic autocrine or paracrine signaling capacity by ISe heifers.

Of the fully annotated cluster 7 DEG, seven genes were downregulated (TNFAIP1L2, CDRT4, PLCZ1, MCART6, FRAS1, SLC35D2, and OR2AT 4) by Mix, three were upregulated (LDHB, VAMP5, and LRRTM2) and one downregulated (PLCZ1) by OSe, and one upregulated (LRRTM2) and one downregulated (SLC35D2) by ISe treatments. Interestingly, when a gene was sensitive to a Se treatment, the response was the same regardless of Se treatment (e.g., downregulation of PLCZ1 by Mix and OSe and upregulation of LRRTM2 by OSe and ISe). Moreover, the effect of Mix versus OSe on gene expression was most different. Genes in cluster 7 are predominately involved with cellular function (LDHB, LRRTM2, PLCZ1, TNFAIP8L2, VAMP5, and FRAS1) and immune system (SLC35D2, TNFAIP8L2) networks.

Of cluster 7 genes, the role of SLC35D2 and TNFAIP8L2, VAMP5 and LLRTM2, and LDHB are well characterized for their roles in immune function, cellular signaling, and energy metabolism, respectively. Specifically, SLC35D2 (HFRC1) is a UMP/UDP-GlcNAc (N-acetylglucosamine) antiporter that is strongly expressed in liver tissue [57] and functions in the Golgi apparatus to transport UDP-GlcNAc from the cytosol into the lumen of the Golgi apparatus. Subsequently, UDP-GlcNAc can be incorporated into oligo/polysaccharide chains [58]. SLC35D2 function also is critical for nucleotide sugar-promoted cell signaling, due to its role in loading of Golgi vesicles with UDP-GlcNAc and its subsequent constitutive pathway-mediated release at the plasma membrane by exocytosis [57]. Accordingly, the decreased expression of SLC35D2 in mix heifers ostensibly suggests a decreased

capacity for UDP-GlcNA-based protein glycosylation. Moreover, decreased SLC35D2 may impart a decreased sensitivity for innate inflammation because of a decreased paracrine signaling of hepatic macrophage recruitment, resulting from a decreased UDP-GlcNAc substrate for binding of macrophage GPR105/P2Y14 [59].

TNFAIP8L2 is a cytoplasma-localized protein that is ubiquitously expressed, including T cells, B cells, macrophages, lymphocytes, glucocorticoid-secreting cells of the adrenal medulla, pancreatic islets, testis, and stomach parietal cells of the stomach [60, 61]. In addition, TNFAIP8L2 is expressed by human, but not mouse, hepatocytes [61]. This widespread pattern of expression indicates that TNFAIP8L2 affects both immune system and non-immune system functions. The finding that TNFAIP8L2 (TIPE2) mRNA was lower in the liver tissue of Mix treatment heifers suggests that TNFAIP8P2, like SLC35D2, is involved in the innate immune system response. Specifically, TNFAIP8L2 is a member of the tumor necrosis factor-alpha-induced protein 8 family and functions to maintain immune homeostasis by negatively regulating Toll-like- and T-cell-receptor-mediated signaling, thus acting to prevent hyper immune responses [62, 63]. In addition, with regard to a non-immune system function of TNFAIP8L2 in the liver, TNFAIP8L2 is thought to be a negative regulator of MAPK and NF- κ B through its association with caspase-8 [61] and caspase-8 is known to be involved with selective apoptotic events in liver tissue cell types [64]. Whether the downregulation of TNFAIP8L2 represents a positive or negative effect on hepatic function remains to be determined and likely depends on the relative abundance of TNFAIP8L2 protein that is expressed [65].

Vesicle-mediated intracellular trafficking is critical for localization of proteins and for secretory and endocytotic pathway-mediated release of proteins and cellular metabolites. Interorganelle vesicle trafficking is dependent on vesicular-SNARE proteins that bind Golgi vesicles and the target-SNARE proteins that bind target membranes. VAMP5 is a member of the vesicle-associated membrane protein family and is expressed by a number of secretory and non-secretory cell types, including the liver [66]. It is thought that VAMP5 functions as a vesicular-SNARE to bind post-Golgi vesicles and target-SNAREs [66, 67], but not two ubiquitous SNAREs synaptin-1 or SNAP25 [67]. The increased mRNA VAMP5 content in OSe heifer livers suggests an increased capacity for secretory and endocytic events.

LRRTM2 is a member of the leucine-rich transmembrane protein family that is expressed by excitatory neurons, localized to postsynaptic membranes, and functions to bind neurexins localized to the presynaptic membrane of adjacent neurons [68, 69]. Thus, LRRTM2 is a critical adhesion molecule in synapses of excitatory (i.e., glutamatergic) neurons. Of the four LRRTM family members, LRRTM2 expression was most widespread outside of brain tissue, with relatively

significant expression by all tissues tested, except for bone marrow, including relatively moderate expression by liver tissue [70]. Upregulation of LRRTM2 is thought to result in increased synapse formation, at least in *in vitro* [68] and *in situ* models [71]. Therefore, that LRRTM2 was upregulated in the livers of OSe treatment heifers suggests that glutamatergic neuron-signaling capacity of the liver was enhanced in these animals.

Lactate dehydrogenase (LDH) is a functional tetramer of LDHA and/or LDHB subunits that mediates the reversible conversion of lactate to pyruvate. The direction of LDH activity depends on the relative composition of LDHA and LDHB subunits; because LDHB has a lower affinity for pyruvate than does LDHA, LDH tetramers with a greater proportion of LDHB subunits results in the conversion of lactate into pyruvate [72]. Although LDHB is ubiquitously expressed, its expression is especially high in the glucose-exporting tissues, the liver and kidney. Therefore, the finding that LDHB mRNA expression in the liver of OSe heifers suggests that these animals had a greater capacity to use lactate for glucose production and/or increased capacity for energy production by its oxidation of lactate.

The effect of Se treatment on cluster 8 genes was similar to that of cluster 7 genes. OSe increased expression of 5 of 20 genes in cluster 8, whereas ISe (five), ISe and Mix (six), or Mix alone (four) decreased expression of DEG. Cluster 8 DEG are involved in cell signaling, molecular and mineral transport, and tissue growth and function. Cluster 8 also contained two of the most strongly affected DEG in the trial (BOLA-DQA2 and CEP350). *B. taurus* major histocompatibility complex, class II, DQ alpha 2 (BOLA-DQA2 [73]) mRNA expression was increased 2.67-fold by OSe, but was insensitive to ISe or Mix treatments. This finding suggests that OSe-consuming animals may have a greater capacity to respond to pathogen challenges [74, 75], assuming that OSe heifers were not stimulated by a pathogen challenge that was not present to the other treatment groups (unlikely given that heifers of different Se treatment groups were housed in the same pens). The additional understanding that Control (0.08 mg Se/day from diet) and Mix (diet plus 1.5 mg supplemental OSe) treatments failed to stimulate BOLA-DQA2 expression suggests that a minimal threshold amount of more than 1.58 mg Se in organic forms is required to stimulate an increased amount of BOLA-DQA mRNA.

In contrast to the increased expression of BOLA-DQA2 by OSe treatment, expression of CEP350 (centromere protein 350 kDa) was decreased 1.5- and 1.39-fold by ISe and Mix, respectively. CEP350 is localized to the centrosome and thought to be involved with the regulation of nuclear hormone receptor function [76]. However, the physiological relevance of decreased CEP350 expression is not obvious.

Neuropeptide Y is one of the most abundant neuropeptides in the mammalian nervous system and affects psychomotor

activity, food intake, regulation of central endocrine secretion, and vasoactivity. The receptor for neuropeptide Y (NPY1R) is expressed by the liver and is associated with control of hepatic blood flow and metabolite secretion [77, 78]. Interestingly, aspartate aminotransferase and alanine aminotransferase are hepatic blood-borne biomarkers of hepatic function and their relative blood-borne activity is associated with NPY1R copy number variation in humans [79]. That NPY1R was downregulated (1.32-fold) in the livers of ISe heifers, but unaffected by Mix or OSe treatments, suggests a decreased NPY1R-mediated regulation of blood flow and metabolite secretion in liver of ISe heifers.

Differentially Expressed miRNA

The GeneChip used for this study can detect genes (MiR) for 621 miRNA. Although all were detected, only the expression of MiR222, MiR2393, and MiR2300b was affected by Se supplementation treatment (Table 1). Whereas the expression of MiR222 has been reported for other species, the expression of MiR2393 and MiR2300b is unique to bovines.

To identify potential physiological consequences of Se treatment alteration of miRNA expression, bioinformatic analysis of predicted mRNA targets of each differentially expressed miRNA was conducted (Table 2). Among the 86 fully annotated DEG, one putative mRNA target was identified for MiR222, 20 for MiR2393, and nine for MiR2300b. Among these mRNA targets, DDHD1 is common to both MiR222 and MiR2393, whereas RUFY3, LRRTM2, and FRAS1 are common targets of both MiR2393 and MiR2300b.

In terms of the potential effect of the three Se treatment-sensitive miRNA on gene network expression, MiR-222 was predicted to affect only one network (Table 2). In contrast, 17 of the 20 differentially expressed targets of MiR-2393 were associated with six networks (the most prominent (nine mRNA) being cell cycle, cancer, and neurological disease) and, for MiR2300b, seven of nine putative target mRNA were identified with four networks (the most prominent (four mRNA) being cell cycle, cellular assembly and organization, DNA repair replication, recombination, and repair).

With regard to actual Se treatment-induced relationship between differentially expressed mRNA and miRNA, MiR222 (cluster 6) was upregulated (1.23-fold) by ISe treatment. Because upregulation of MiR likely is associated with decreased gene expression, this result suggests that ISe treatment has a negative effect on genes controlled by MiR222. Of the 85 fully annotated DEG, only seven (GCG, cluster 3; NPY1R, TMEM168, THG1L, THOC5, ADNP2, cluster 8) shared this pattern of ISe-induced decreased expression (Table 1). However, none of these are putative targets of MiR222. Instead, only DDHD1 is a putative target (Table 2). However, DDHD1 expression was upregulated 1.30-fold by Mix treatment (Table 1). Thus, (a) DDHD1 was not a target of

MiR222 and (b) the consequence of MiR222 upregulation by Mix Se treatment is not obvious.

In contrast to MiR222, MiR2393 expression was decreased by Se treatment. Specifically, MiR2393 (cluster 4) expression was downregulated by Mix (1.41-fold) and OSe (1.65-fold), and tended to be decreased (1.20-fold) by ISe (Table 1). A decrease in MiR likely would result in an increased expression of its targets. However, of the 17 putative targets of MiR2393 (Table 2), only one of these (QPCT) was in cluster 4, and any form of Se supplement downregulated QPCT expression (Table 1). Therefore, QPCT does not appear to be a target of MiR2393. In contrast, DDHD1 (cluster 3) was upregulated by Mix, concomitant with downregulation of MiR2393 by Mix treatment (Table 1). Thus, the data are consistent with DDHD1 expression being regulated by Mix treatment down-regulation of MiR2393. DDHD1 can generate the arachidonic acid-containing lysophosphatidylinositol that increases both the phosphorylation of the extracellular signal-regulated kinase (ERK) and intracellular Ca^{2+} levels [80]. TRP5C also is a putative target of MiR2393 whose expression was upregulated by ISe (1.37-fold), Mix (1.58-fold), and OSe (1.67-fold). As noted before, TRP5C (cluster 1) expression increases the capacity for Ca^{2+} influx into cells [24]. Together, the concomitant downregulation of MiR2393 and upregulation of DDHD1 and TRP5C by at least Mix and OSe supplementation suggests one mechanism by which Se supplementation of Se-inadequate diets results in increased redox potential [25]. This supposition is strengthened by the observation that SEPW1 (cluster 1), which functions as an antioxidant enzyme [23], also is upregulated by all three Se supplements (Table 1).

The effect of Se supplementation on liver expression of MiR2300b (cluster 8), was to decrease it (Table 1). Thus, the expression of MiR2300b targets should be increased. Unlike MiR2393, MiR2300b was downregulated by ISe and Mix treatments and unaffected by OSe. Of the putative seven DEG MiR2300b targets (Table 2), four (FRAS1, RUFY3, SNAPIN, and THG1L) were downregulated by Se treatment and thus not likely regulated by MiR2300b expression. Of the remaining three putative targets, LRRTM2 was upregulated by ISe and OSe, SNHG3-RCC1 was upregulated by Mix, and SEPW1 was upregulated by all three Se supplement forms. Because the expression patterns of these putative targets of MiR2300b do correlate with that of MiR2300b, it is unlikely that their expression was regulated by MiR2300b.

Comparison of Current to Previous Studies on the Effect of Dietary Se Form on Hepatic Gene Expression in Developing Beef Heifers

Previously, we found that dietary supplementation of none or 3 mg Se/day as ISe or OSe to maturing beef heifers [9] resulted in assimilation of Se by the liver that was essentially found in the present study for Control, ISe, and OSe heifers. In contrast

Table 2 Hepatic miRNA (Mir) and putative mRNA targets differentially expressed in response to form of dietary selenium supplement

miRNA	Molecules in network ^a	Top functions ^b
MiR222	<u>ADCYAP1</u> , <u>DDHD1</u>	Cell morphology, cell-to-cell signaling and interaction, cellular assembly and organization
MiR2393	<u>ABCF3</u> , <u>ACIN1</u> , <u>AKR1C3</u> , <u>CASP2</u> , <u>CBX1</u> , <u>CBX3</u> , <u>CCNA1</u> , <u>CCND1</u> , <u>CDKN2A</u> , <u>CLDN2</u> , <u>CTNNB1</u> , <u>DIABLO</u> , <u>GCM1</u> , <u>GOLGA3</u> , <u>GSK3B</u> , <u>HDAC1</u> , <u>HIC1</u> , <u>HNF1A</u> , <u>HNF4A</u> , <u>HTRA2</u> , <u>LRRTM2</u> , <u>MF12</u> , <u>NFYC</u> , <u>PPARA</u> , <u>QPCT</u> , <u>RBM39</u> , <u>RIN2</u> , <u>RUFY3</u> , <u>SAP18</u> , <u>SENP1</u> , <u>SRPK2</u> , <u>SRR</u> , <u>TANK</u> , <u>TSNAX</u> , <u>ZNF236</u> <u>ADA</u> , <u>ADCYAP1</u> , <u>anandamide</u> , <u>Ca²⁺</u> , <u>CCL22</u> , <u>chemokine</u> , <u>CIB1</u> , <u>CXCL2</u> , <u>CXCR2</u> , <u>D-sphingosine</u> , <u>DDHD1</u> , <u>ENTPD1</u> , <u>Eotaxin</u> , <u>FCGR2A</u> , <u>FKBP4</u> , <u>HOMER1</u> , <u>Integrin alpha 4 beta 1</u> , <u>ITGA2</u> , <u>ITPR3</u> , <u>LAMA1</u> , <u>leukotriene B4</u> , <u>Lfa-1</u> , <u>MOG</u> , <u>Plcbeta</u> , <u>PTAFR</u> , <u>Rap1</u> , <u>SELPLG</u> , <u>SFTP</u> , <u>SLC9A3R1</u> , <u>TRP</u> , <u>TRPC3</u> , <u>TRPC4</u> , <u>TRPC5</u> , <u>VIPR1</u> , <u>Vla-4</u> <u>CRK</u> , <u>PCDHA2</u>	Cell cycle, cancer, neurological disease
	<u>E2F1</u> , <u>ZNF22</u>	Cell signaling, molecular transport, vitamin and mineral metabolism
	<u>DLG4</u> , <u>FRAS1</u> , <u>GRIP1</u> , <u>GRIP2</u>	
	<u>ADNP2</u> , <u>CBX1</u> , <u>CBX3</u> , <u>NFYC</u>	
MiR2300b	<u>BCAR3</u> , <u>CCND1</u> , <u>CDK4</u> , <u>CSNK1A1</u> , <u>CSNK1D</u> , <u>CSNK1E</u> , <u>CTNNB1</u> , <u>DDR1</u> , <u>dihydrotestosterone</u> , <u>DYNC1H1</u> , <u>EBAG9</u> , <u>HADHB</u> , <u>HIST2H2BE</u> , <u>KAT7</u> , <u>KPNA4</u> , <u>LRRTM2</u> , <u>MAPK14</u> , <u>NDC80</u> , <u>NGFR</u> , <u>NPY</u> , <u>PCM1</u> , <u>PKM2</u> , <u>RALY</u> , <u>RANBP1</u> , <u>RCC1</u> , <u>RGS7</u> , <u>RUFY3</u> , <u>SNAPIN</u> , <u>TP53</u> , <u>TRPV1</u> , <u>Tubulin</u> , <u>UNC119</u> , <u>VRK1</u> , <u>XPO1</u> , <u>ZFP106</u> <u>ELMOD1</u> , <u>THG1L</u>	Embryonic development, organ development, organismal development
	<u>DLG4</u> , <u>FRAS1</u> , <u>GRIP1</u> , <u>GRIP2</u>	Cancer, cardiac dysfunction, cell cycle
	<u>CTNNB1</u> , <u>dihydrotestosterone</u> , <u>SECISBP2</u> , <u>SEPW1</u>	Developmental disorder, genetic disorder, ophthalmic disease
		DNA replication, recombination, and repair; gene expression, cellular assembly and organization
		Cell cycle, cellular assembly and organization, DNA replication, recombination, and repair
		Tissue morphology, cellular growth and proliferation, cell death
		Developmental disorder, genetic disorder, ophthalmic disease
		Organismal injury and abnormalities, renal damage, renal and urological disease

Differences in expression and dietary Se treatments are described in Table 1

^aThe mRNA for the underlined proteins were differentially expressed by the liver in response to supplemental Se treatment

^bThe top biofunctions are identified for the differentially expressed genes in the left

to the present study, only 80 DEG were identified and only one (LEPRE1) was commonly altered in both trials. In addition, LEPRE1 was upregulated by both ISe (1.53-fold) and OSe (1.47-fold) treatments (Mix treatment was not tested), whereas in the present trial, LEPRE1 was not affected by ISe or OSe treatment but was downregulated (1.15-fold) by Mix treatment.

Although it may seem surprising that more similarities were not found between the two studies, given that the same experimental model was used and very similar hepatic Se contents was achieved for the comparable Control, ISe, and OSe treatments between the two trials, the lack of similarity between gene expression patterns found for the two trials may be reasonable given the following experimental differences: (1) although similar in magnitude, Se concentrations in liver after 105/106 days of supplementation [9] may not have been stable, as they were after 168 days (current trial); (2) different experimental designs were used (non-supplemented Control, ISe, and OSe and $n=6$ /treatment for Liao et al. [9] versus non-

supplemented Control, ISe, OSe, and Mix and $n=9$ /treatment for the present study); (3) differences in the microarray assays to detect gene expression due to the use of different GeneChips; and (4) a combination of the above. With regard to the third possibility, Liao et al. [9] used the commercially available Affymetrix GeneChip Bovine Genome Array, which contains 24,016 gene transcripts and which was annotated using the NetAffix annotation database for 3' IVT Expression on Bovine GeneChip Array 3'-weighted coverage of transcripts (<http://www.affymetrix.com/analysis/index:affx>). In contrast, the present trial used the custom WT Btau 4.0 Array (version 1), which contained 26,303 gene transcripts, a complete 5' to 3' coverage of gene transcripts, and annotation based on an updated draft (4.0) of the Bovine Genome. In addition, the Btau 4.0 GeneChip used in the present study contained 621 microRNA, whereas the Affymetrix GeneChip Bovine Genome Array used in the previous study [9] contained none.

Form of Se-Induced Physiological Phenotypes

To facilitate comparison of form of Se-induced alterations in physiological capacities, due to consumption of OSe, Mix, or ISe relative to Control heifers, DEG, and their putative altered physiological functions described above were grouped by Se treatment without regard for their hierarchical clusters (Table 3). Relative to Se-inadequate nonsupplemented Control Se treatment, supplementation of 3 mg Se/day by OSe, Mix, and ISe resulted in common upregulation of antioxidant (SEPW1), sensing of oxidative metabolites (TRPC5) and IL-3-mediated immune response (ITGA2), and decreased chemokine-mediated inflammation (QPCT). In addition,

albeit by altering expression of different genes, a common effect of Se supplementation was an increased glutamatergic neuron-mediated signaling (LRRTM2, OSe, ISe; SLC6A17, Mix, and ISe) potential.

In addition to these common effects of Se supplementation on hepatic gene transcription profiles (relative to the nonsupplemental Control), the OSe “phenotype” included decreased Rab5-specific endocytic signaling but increased total endocytic pathway-mediated secretion, increased immune system capacity, and increased use of lactate. In contrast, the Mix “phenotype” uniquely included increased capacity for energy substrate (fatty acids, glucose) absorption and oxidation and (consequently,

Table 3 Putative differences in physiological functions of liver tissue in maturing beef heifers fed organic (OSe), inorganic (ISe), or a 1:1 mix (Mix) of OSe and ISe forms of supplemental selenium (Se) relative to nonsupplemented heifers

Gene(s)	ISe	OSe	Mix	Putative altered physiological capacity ^a
Common response				
SEPW1	Up	Up	Up	Increased antioxidant capacity
TRPC5	Up	Up	up	Increased sensing of oxidative metabolites
ITGA2	Up	Up	Up	Increased IL-3-mediated immune response and wound healing
QPCT	Down	Down	Down	Decreased chemokine-mediated inflammation
Differential response				
CYP2D14		Down		Decreased alkaloid and xenobiotic catabolism capacity
RUFY3		Down		Increased neuronal polarity by decreased excess axon formation
RIN2		Down		Decreased Rab5-mediated endocytotic signaling
GSG2		Down		Increased regulation of chromosome segregation
ZNF238		Down		Increased capacity for transcription
MiRNA 2393		Down	Down	Increased expression of miRNA 2393 target mRNA
BOLA-DQ2a		Up		Increased adaptive immune response
VAMP5		UP		Increased capacity for endocytotic pathway-mediated secretion
LDHB		Up		Increased use of lactate for glucose and/or ATP production
LRRTM2	Up	Up		Increased capacity for glutamatergic neuron signaling
MiR2300b	Down		Down	Increased expression of miR2300b target mRNA
PCDHB14	Up		Up	Increased regulation of hepatic neuronal development
DDHD1			Up	Increased capacity for fatty acid and glucose absorption and oxidation
eIf4			Up	Increased capacity for protein translation
PSME4			Up	Increased capacity to respond to metabolism-induced oxidation [81–85]
CCNB2			Up	Increased capacity for hepatic regeneration [86, 87]
GJB4	Up		Up	Increased immune system function and increased inter-hepatocyte signaling [88–90]
SLC6A17	Up		Up	Increased glutamatergic neuron-mediated signaling
TNFAIP8L2			Down	Increased MAPK- and NFκβ-mediated signaling
SLC35D2	Down		Down	Decreased sensitivity for innate immune system-induced inflammation
CEP350	Down		Down	Unknown
SNAPIN	Down		Down	Decreased synaptic transmission capacity [90, 91]
GHRH	Up			Increased autocrine or paracrine signaling
MiR222	Up			Decreased expression of miR222 targets mRNA
GCG	Down			Decreased glucagon-mediated signaling
NPY1R	Down			Decreased neuropeptide Y-mediated signaling

^a References are provided for those physiological capacities not described in the “Results and Discussion” section

perhaps) protein translation. In addition, the Mix phenotype also had the capacity for increased MAPK- and NFκβ-mediated signaling, seemingly consistent with increased putative capacity for increased interhepatocyte signaling and/or immune system function. In contrast to the OSe- and Mix-induced DEG profiles, a striking feature of the ISe “phenotype” was the decreased potential for regulation of hepatic capacities by glucagon and neuropeptide-Y regulation but an increased apparent sensitivity for regulation by GHRH-mediated paracrine and/or autocrine signaling.

Summary

The objectives of the current work were (a) to compare hepatic gene transcriptome profiles when hepatic Se assimilation was stable (day 168) using microarray methods and (b) to conduct bioinformatic analysis of these profiles to predict altered hepatic physiological capacities induced by specific forms of Se supplement, contained in a common basal mineral mix. Hepatic gene transcriptome profiles were successfully compared by microarray analysis, resulting in the understanding that the chemical form (inorganic, organic, 1:1 mix of inorganic and organic) of supplemental Se differentially affected hepatic gene expression of growing heifers after 168 days of supplementation. The effects of Se treatments on the global pattern of gene expression appeared to be more similar for Control and OSe than Control and ISe, and Mix was between OSe and ISe. This finding suggests that the different hepatic gene expression profile of ISe, Mix, OSe, and Control heifers, was not caused by an interform gradient (0, 50, 100 % ISe or OSe) effect, but by their biochemical form and form-specific metabolic pathways. Bioinformatic analysis indicated that Se treatment-induced differentially expressed hepatic genes were indicative of altered capacity for physiological challenges, including increased capacity for redox potential induced by OSe and Mix Se treatments resulting from decreased expression of MiR2300b mRNA. Moreover, collectively, bioinformatic analyses revealed distinct OSe-, Mix-, and ISe-Se treatment-induced “phenotypes” that possessed both common and unique predicted physiological capacities.

Acknowledgments This research was supported by the Alltech-University of Kentucky Animal Nutrigenomics Alliance (JCM), University of Kentucky, and Kentucky Agricultural Experiment Station. The information reported in this paper (publication no. 14-07-038) is part of a project of the Kentucky Agricultural Experiment Station and is published with approval of the Director.

References

- Suzuki KT (2005) Metabolomics of selenium: Se metabolites based on speciation studies. *J Health Sci* 51:107–114
- Pehrson B (1993) Diseases and diffuse disorders related to selenium deficiencies in ruminants. *Nor J Agric Sci* 11:79–93
- Sordillo LM, Aitken SL (2009) Impact of oxidative stress on the health and immune function of dairy cattle. *Vet Immunol Immunopathol* 128:104–109
- National Research Council (1996) Minerals. In: Nutrient requirements of beef cattle, 7th revised edition. 2001 Update. National Academy Press, Washington, DC, pp 54–74
- Ammerman CB, Miller SM (1975) Selenium in ruminant nutrition: a review. *J Dairy Sci* 58:1561–1571
- Dargatz DA, Ross PF (1996) Blood selenium concentrations in cows and heifers on 253 cow-calf operations in 18 states. *J Anim Sci* 74: 2891–2895
- Food and Drug Administration (1987) Food additives permitted in feed and drinking water of animals: selenium. *Fed Reg* 52:10668
- Korhola M, Vainio A, Edelmann K (1986) Selenium yeast. *Ann Clin Res* 18:65–68
- Liao SF, Brown KR, Stromberg AJ, Burris WR, Boling JA, Matthews JC (2011) Dietary supplementation of selenium in inorganic and organic forms differentially and commonly alters blood and liver selenium concentrations and liver gene expression profiles of growing beef heifers. *Biol Trace Elem Res* 140:151–169
- Brennan KM, Burris WR, Boling JA, Matthews JC (2011) Selenium content in blood fractions and live of beef heifers is greater with a mix of inorganic/organic or organic versus inorganic supplemental selenium but the time required for maximal assimilation is tissue-specific. *Biol Trace Elem Res* 144:504–516
- Ortman K, Pehrson B (1999) Effect of selenate as a feed supplement to dairy cows in comparison to selenite and selenium yeast. *J Anim Sci* 77:3365–3370
- Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 4:249–264
- Wu Z, Irizarry RA, Gentleman R, Murillo FM, Spencer F (2004) A model based background adjustment for oligonucleotide expression arrays. *J Am Stat Assoc* 99:909–917
- Partek D (2009) Partek documentation: turning data into discovery. Partek Incorporated, St. Louis
- Lewis BP, Burge CP, Bartel DP (2005) Conserved seed pairing, often flanked by adenosine, indicates that thousands of human genes are microRNA targets. *Cell* 120:15–20
- Friedman RC, Farh KK, Burge CB, Bartel DP (2009) Most mammals are conserved targets of microRNAs. *Genome Res* 19:92–105
- Fischer A, Pallauf J, Gohil K, Weber SU, Packer L, Rimbach G (2001) Effect of selenium and vitamin E deficiency on differential gene expression in rat liver. *Biochem Biophys Res Commun* 285: 470–475
- Huang J-Q, Li D-L, Zhao H, Sun L-H, Xia X-J, Wang K-N, Luo X, Lei XG (2011) The selenium deficiency disease exudative diathesis in chicks is associated with down regulation of seven common selenoprotein genes in liver and muscle. *J Nutr* 141: 1605–1610
- Sun B, Wang RH, Li J, Jiang Z, Xu S (2011) Dietary selenium affects selenoprotein W gene expression in the liver of chicken. *Biol Trace Elem Res* 143:1516–1523
- Ullrey DE (1987) Biochemical and physiological indicators of selenium status in animals. *J Anim Sci* 65:1712–1726
- Corah LR (1996) Trace mineral requirements of grazing cattle. *Anim Feed Sci Technol* 59:61–70

22. Surai PF (2006) Selenium in ruminant nutrition. In: Selenium in nutrition and health. Nottingham University Press, Nottingham, UK, pp 487–587
23. Jeong D, Kim TS, Chung YW, Lee BJ, Kim IY (2002) Selenoprotein W is a glutathione-dependent antioxidant in vivo. *FEBS Lett* 517: 225–228
24. Mori Y, Kajimoto T, Nakao A, Takahashi N, Kiyonaka S (2011) Receptor signaling integration by TRP channelsomes. *Adv Exp Med Biol* 704:373–389
25. Yamamoto S, Takahashi N, Mori Y (2010) Chemical physiology of oxidative stress-activated TRPM2 and TRPC5 channels. *Prog Biophys Mol Biol* 103:18–27
26. Bon RS, Beech DJ (2013) In pursuit of small molecule chemistry for calcium-permeable non-selective TRPC channels – mirage or pot of gold? *Br J Pharmacol* 170:459–474
27. Riccio A, Medhurst AD, Mattei C, Kelsell RE, Calver AR, Randall AD, Benham CD, Pangalos MN (2002) mRNA distribution analysis of human TRPC family in CNS and peripheral tissues. *Brain Res Mol Brain Res* 109:95–104
28. Xu SZ, Sukumar P, Zeng F, Li J, Jairaman A, English A, Naylor J, Ciurtoni C, Majeed Y, Milligan CJ, Bahnsi YM, Al-Shawaf E, Porter KE, Jiang LH, Emery P, Sivaprasadarao A, Beech DJ (2008) TRPC channel activation by extracellular thioredoxin. *Nature* 451:69–72
29. Rychkov GY, Barritt G (2011) Expression and function of TRP channels in liver cells. *Adv Exp Med Biol* 704:667–686
30. Guo H, Nairn A, dela Rosa M, Nagy T, Zhao S, Moreman K, Pierce M (2012) Transcriptional regulation of the protocadherin β lucster during Her-2 protein-induced mammary tumorigenesis results from altered N-glycan branching. *J Biol Chem* 287:24941–24954
31. Kitagishi Y, Matsuda S (2013) RUFY, Rab and Rap family proteins involved in a regulation of cell polarity and membrane trafficking. *Int J Mol Sci* 14:6487–6498
32. Nakazawa H, Sada T, Toriyama M, Tago K, Sugiura T, Fukuda M, Inagaki N (2012) Rab33a mediates anterograde vesicular transport for membrane exocytosis and axon outgrowth. *J Neurosci* 32:12712–12725
33. Tsuneoka Y, Matsuo Y, Higuchi R, Ichikawa Y (1992) Characterization of the cytochrome P-450IID subfamily in bovine liver. Nucleotide sequences and microheterogeneity. *Eur J Biochem* 280:739–746
34. Zanger UM, Schwab M (2013) Cytochrome P450 enzymes in drug metabolism: regulation of gene expression, enzyme activities, and impact of genetic variation. *Pharmacol Ther* 138:103–111
35. Ingelman-Sundberg M (2005) Genetic polymorphisms of cytochrome P450 2D6 (CYP2D6): clinical consequences, evolutionary aspects and functional diversity. *Pharmacogen J* 5:6–13
36. Kimura T, Sakisaka T, Baba T, Yamada T, Takai Y (2006) Involvement of the Ras-Ras-activated Rab5 guanine nucleotide exchange factor RIN2-Rab5 pathway in the hepatocyte growth factor-induced endocytosis of E-cadherin. *J Biol Chem* 281:10598–10609
37. Zeigerer A, Gilleron J, Bogorad RL, Marsico G, Nonaka H, Seifert S, Epstein-Barash H, Kuchimanchi S, Peng CG, Ruda VM, Del Conte-Zerial P, Hengstler JG, Kalaidzidis Y, Koteliansky V, Zerial M (2012) Rab5 is necessary for the biogenesis of the endolysosomal system in vivo. *Nature* 485:465–470
38. Yamashita A, Oka S, Tanikawa T, Hayashi Y, Neomoto-Sasaki Y, Sugiura T (2013) The actions and metabolism of lysophosphatidylinositol, an endogenous agonist for GPR55. *Prostagland Lipid Mediat* 107:103–116
39. Higgs HN, Han MH, Johnson GE, Glomset JA (1998) Cloning of a phosphatidic acid-preferring phospholipase A1 from bovine testis. *J Biol Chem* 273:5468–5477
40. Moreno-Navarrete JM, Catalán V, Whyte L, Díaz-Arteaga A, Vázquez-Martínez R, Rotellar F, Guzmán R, Gómez-Ambrosi J, Pulido MR, Russell WR, Imbernón M, Ross RA, Malagón MM, Dieguez C, Fernández-Real JM, Frühbeck G, Nogueiras R (2012) The l-alpha-lysophosphatidylinositol/GPR55 system and its potential role in obesity. *Diabetes* 61:281–291
41. Simocks AC, O'Keefe L, Jenkin KA, Mathai ML, Hryciw DH, McAinch AJ (2011) A potential role for GPR55 in the regulation of energy homeostasis. *Drug Discov Today*. doi:10.1016/j.drudis.2013.12.005
42. Svitkin YV, Pause A, Haghigiat A, Pyronnet S, Witherell G, Belsham GJ, Sonenberg N (2001) The requirement for eukaryotic initiation factor 4A (eIF4A) in translation is in direct proportion to the degree of mRNA 5' secondary structure. *RNA* 7:382–394
43. Wang FW, Ulyanova NP, van der Waal MS, Patnaik D, Lens SM, Higgins JM (2011) A positive feedback loop involving Haspin and Aurora B promotes CPC accumulation at centromeres in mitosis. *Curr Biol* 21:1061–1069
44. Okado H, Ohata-Karayama C, Sugitani Y, Fukuda Y, Ishida R, Hirai S, Miwa A, Takahashi A, Aoki K, Mochida K, Suzuki O, Honda T, Nakajima K, Ogawa M, Terashima T, Matsuda J, Kawano H, Kasai M (2009) The transcriptional repressor RP58 is crucial for cell-division patterning and neuroanl survival in the developing cortex. *Dev Biol* 331:140–151
45. Kelly KF, Daniel JM (2006) POZ for effect-POZ-ZF transcription factors in cancer and development. *Trends Cell Biol* 16:578–587
46. Fischer WH, Spiess J (1987) Identification of a mammalian glutaminyl cyclase converting glutaminyl into pyroglutamyl peptides. *PNAS USA* 84:3628–3632
47. Cynis H, Kehlen A, Haegele M, Hoffmann T, Fujii U, Shibasaki Y, Yoneyama H, Schilling S, Demuth HU (2013) Inhibition of glutaminyl cyclases alleviates CCL2-mediated inflammation of non-alcoholic fatty liver disease in mice. *Int J Exp Pathol* 94:217–225
48. Friedman SL (2000) Molecular regulation of hepatic fibrosis, an integrated cellular response to tissue injury. *J Biol Chem* 275:2247–2250
49. Yakubenko VP, Bhattacharjee A, Pluskota E, Cathcart MK (2011) alphaMbeta2 Integrin activation prevents alternative activation of human and murine macrophages and impedes foam cell formation. *Circ Res* 108:544–554
50. Zaia KA, Reimer RJ (2009) Synaptic vesicle protein NTT4/XT1 (SLC6A17) catalyzes Na^+ -coupled neutral amino acid transport. *J Biol Chem* 284:8439–8448
51. Hohla F, Moder A, Mayrhoiser U, Hauser-Kronberger C, Schally AV, Varga JL, Zarandi M, Buchholz S, Huber R, Aigner E, Ritter M, Datz C (2008) Differential expression of GHRH receptor and its splice variant 1 in human normal and malignant mucosa of the oesophagus and colon. *Int J Oncol* 33:137–143
52. Kiaris H, Chatzistamou I, Papavassiliou AG, Schally AV (2011) Growth hormone-releasing hormone: not only a neurohormone. *Trends Endocrinol Metab* 22:311–317
53. Christodoulou C, Schally AV, Chatzistamou I, Kondi-Pafiti A, Lamnisso K, Kouloheri S, Kalofoutis A, Kiaris H (2006) Expression of growth hormone-releasing hormone (GHRH) and splice variant of GHRH receptors in normal mouse tissues. *Regul Pept* 136:105–108
54. Bagnato A, Moretti C, Ohnishi J, Frajese G, Catt KJ (1992) Expression of the growth hormone-releasing hormone gene and its peptide product in the rat ovary. *Endocrinology* 130:1097–1102
55. Berry SA, Srivastava CH, Rubin LR, Phipps WR, Pescovitz OH (1992) Growth hormone-releasing hormone-like messenger ribonucleic acid and immunoreactive peptide are present in human testis and placenta. *J Clin Endocrinol Metab* 75:281–284
56. Bosman FT, Van Assche C, Nieuwenhuyzen Kruseman AC, Jackson S, Lowry PJ (1984) Growth hormone-releasing factor (GRF) immunoreactivity in human and rat gastrointestinal tract and pancreas. *J Histochem Cytochem* 32:1139–1144
57. Sesma JI, Esther CR Jr, Kreda SM, Jones L, O'Neal W, Nishihara S, Nicholas RA, Lazarowski ER (2009) Endoplasmic reticulum/golgi nucleotide sugar transporters contribute to the cellular release of UDP-sugar signaling molecules. *J Biol Chem* 284:12572–12583

58. Ishida N, Kuba T, Aoki K, Miyatake S, Kawakita M, Sanai Y (2005) Identification and characterization of human Golgi nucleotide sugar transporter SLC35D2, a novel member of the SLC35 nucleotide sugar transporter family. *Genomics* 85:106–116
59. Xu J, Morinaga H, Oh D, Li P, Chen A, Talukdar S, Mamane Y, Mancini JA, Nawrocki AR, Lazarowski E, Olefsky JM, Kim JJ (2012) GPR105 ablation prevents inflammation and improves insulin sensitivity in mice with diet-induced obesity. *J Immunol* 189:1992–1999
60. Zhang G, Hao C, Lou Y, Xi W, Wang X, Wang Y, Qu Z, Guo C, Chen Y, Zhang Y, Liu S (2010) Tissue-specific expression of TIPE2 provides insights into its function. *Mol Immunol* 47:2435–2442
61. Zhang L, Shi Y, Wang Y, Zhu F, Wang Q, Ma C, Chen YH, Zhang L (2011) The unique expression profile of human TIPE2 suggests new functions beyond its role in immune regulation. *Mol Immunol* 48: 1209–1215
62. Sun H, Gong S, Carmody RJ, Hilliard A, Li L, Sun J, Kong L, Xu L, Hilliard B, Hu S, Shen H, Yang X, Chen YH (2008) TIPE2, a negative regulator of innate and adaptive immunity that maintains immune homeostasis. *Cell* 133:415–426
63. Lou Y, Liu S, Zhang C, Zhang G, Li J, Ni M, An G, Dong M, Liu X, Zhu F, Zhang W, Gao F, Chen YH, Zhang Y (2013) Enhanced atherosclerosis in TIPE-2deficient mice is associated with increased macrophage responses to oxidized low-density lipoprotein. *J Immunol* 191:4849–4857
64. Chaudhary K, Liedtke C, Wertenbruch S, Trautwein C, Streetz KL (2013) Caspase8differentially controls hepatocytes and non-parenchymal liver cells during chronic cholestatic liver injury in mice. *J Hepatol* 59:1292–1298
65. Gus-Brautbar Y, Johnson D, Zhang L, Sun H, Wang P, Zhang S, Zhang L, Chen YH (2012) The anti-inflammatory TIPE2 is an inhibitor of the oncogenic Ras. *Mol Cell* 45:610–618
66. Zeng Q, Subramaniam VN, Wong SH, Tang BL, Parton RG, Rea S, James DE, Hong W (1998) A novel synaptobrevin/VAMP homologous protein (VAMP5) is increased during *in vitro* myogenesis and present in the plasma membrane. *Mol Biol Cell* 9:2423–2437
67. Hasan N, Corbin D, Hu C (2010) Fusogenic pairings of vesicle-associated membrane proteins (VAMPs) and plasma membrane t-SNAREs–VAMP5 as the exception. *PLoS ONE* 5:e14238
68. Ko J, Fuccillo MV, Malenka RC, Südh (2009) LRRTM2 functions as a neurexin ligand in promoting excitatory synapse formation. *Neuron* 64:791–798
69. Bang ML, Owczarek S (2013) A matter of balance: role of neurexin and neuroligin at the synapse. *Neurochem Res* 38:1174–1189
70. Laurén J, Airaksinen MS, Saarma M (2003) Timmusk T (2003) A novel gene family encoding leucine-rich repeat transmembrane proteins differentially expressed in the nervous system. *Genomics* 81: 411–421
71. Soler-Llavina GJ, Fuccillo MV, Ko J, Sudhof TC, Malenka RC (2011) The neurexin ligands, neuroligins and leucine-rich repeat transmembrane proteins, perform convergent and divergent synaptic functions *in vivo*. *PNAS* 108:16502–16509
72. Doherty JR, Cleveland JL (2013) Targeting lactate metabolism for cancer therapeutics. *J Clin Invest* 123:3685–3692
73. Morooka A, Asahina M, Kohda C, Tajima S, Niimi M, Nishino Y, Sugiyama M, Aida Y (1995) Nucleotide sequence and the molecular evolution of a new A2 gene in the DQ subregion of the bovine major histocompatibility complex. *Biochem Biophys Res Commun* 212: 110–117
74. Handunnetthi L, Ramagopalan SV, Ebers GC, Knight JC (2009) Regulation of MHC II gene expression, genetic variation, and disease. *Genes Immun* 11:99–112
75. Hou Q, Huang J, Ju Z, Li Q, Li L, Wang C, Sun T, Wang L, Hou M, Hang S, Zhong J (2012) Identification of splice variants, targeted microRNAs and functional single nucleotide polymorphisms of the BOLA-DQ2 gene in dairy cattle. *DNA Cell Biol* 31:739–744
76. Zimin AV, Delcher AL, Florea L, Kelley DR, Schatz MC, Puiu D, Hanrahan F, Pertea G, Van Tassell CP, Sonstegard TS, Marçais G, Roberts M, Subramanian P, Yorke JA, Salzberg SL (2009) A whole genome-assembly of the domestic cow, *Bos taurus*. *Genome Biol* 10(4):R42
77. Ding W, Fujimura M, Mori A, Tooyama I, Kimura H (1991) Light and electron microscopy of neuropeptide Y-containing nerves in human liver, gallbladder, and pancreas. *Gastroenterology* 101: 1054–1059
78. El-Salhy M (2000) Neuropeptide levels in murine liver and biliary pathways. *Ups J Med Sci* 105:207–213
79. Kim HY, Cho S, Yu J, Sung S, Kim H (2010) Analysis of copy number variation in 8,842 Korean individuals reveals 39 genes associated with hepatic biomarkers AST and ALT. *BMB Rep* 43: 547–553
80. Yamashita A, Kumazawa T, Koga H, Suzuki N, Oka S, Sugiura T (2010) Generation of lysophosphatidylinositol by DDHD domain containing 1 (DDHD1): possible involvement of phospholipase D/phosphatidic acid in the activation of DDHD1. *Biochim Biophys Acta* 1801:711–720
81. Savulescu AF, Glickman MH (2011) Proteasome activator 200: The HEAT is on.... *Mol Cell*. doi:10.1074/mcp.M110.006890
82. Ustrell V, Hoffman L, Pratt G, Rechsteiner M (2002) PA200, a nuclear proteasome activator involved in DNA repair. *EMBO J* 21: 3516–3525
83. Khor B, Bredemeyer AL, Huang CY, Turnbull IR, Evans R, Maggi LB Jr, White JM, Walker LM, Carnes K, Hess RA, Sleckman BP (2006) Proteasome activator PA200 is required for normal spermatogenesis. *Mol Cell Biol* 26:2999–3007
84. Davies KJ (2001) Degradation of oxidized proteins by the 20S proteasome. *Biochimie* 83:301–310
85. Jung T, Bader N, Grune T (2007) Oxidized proteins: intracellular distribution and recognition by the proteasome. *Arch Biochem Biophys* 462:231–237
86. Gong D, Ferrell JE Jr (2010) The roles of cyclin A2, B1, and B2 in early and late mitotic events. *Mol Biol Cell* 21:3149–3161
87. Wang X, Quail E, Hung NJ, Tan Y, Ye H, Costa RH (2001) Increased levels of forkhead box M1B transcription factor in transgenic mouse hepatocytes prevent age-related proliferation defects in regenerating liver. *Proc Natl Acad Sci U S A* 98: 11468–11473
88. Alves LA, Coutinho-Silva R, Persechini PM, Spay DC, Savino W, Campos de Carvalho AC (1996) Are there functional gap junctions or junctional hemichannels in macrophages? *Blood* 88:328–334
89. Alves LA, Campos de Carvalho AC, Savino W (1998) Gap junctions: a novel route for direct cell-cell communication in the immune system. *Immunol Today* 19:269–275
90. Ilardi JM, Mochida S, Sheng ZH (1999) Snapin: a SNARE-associated protein implicated in synaptic transmission. *Nat Neurosci* 2:119–124
91. Pan PY, Tian JH, Sheng ZH (2009) Snapin facilitates the synchronization of synaptic vesicle fusion. *Neuron* 61:412–424

Hematologic Abnormalities During the Use of Low Dose Methotrexate for Rheumatoid Arthritis: A Systematic Review and Meta-Analysis

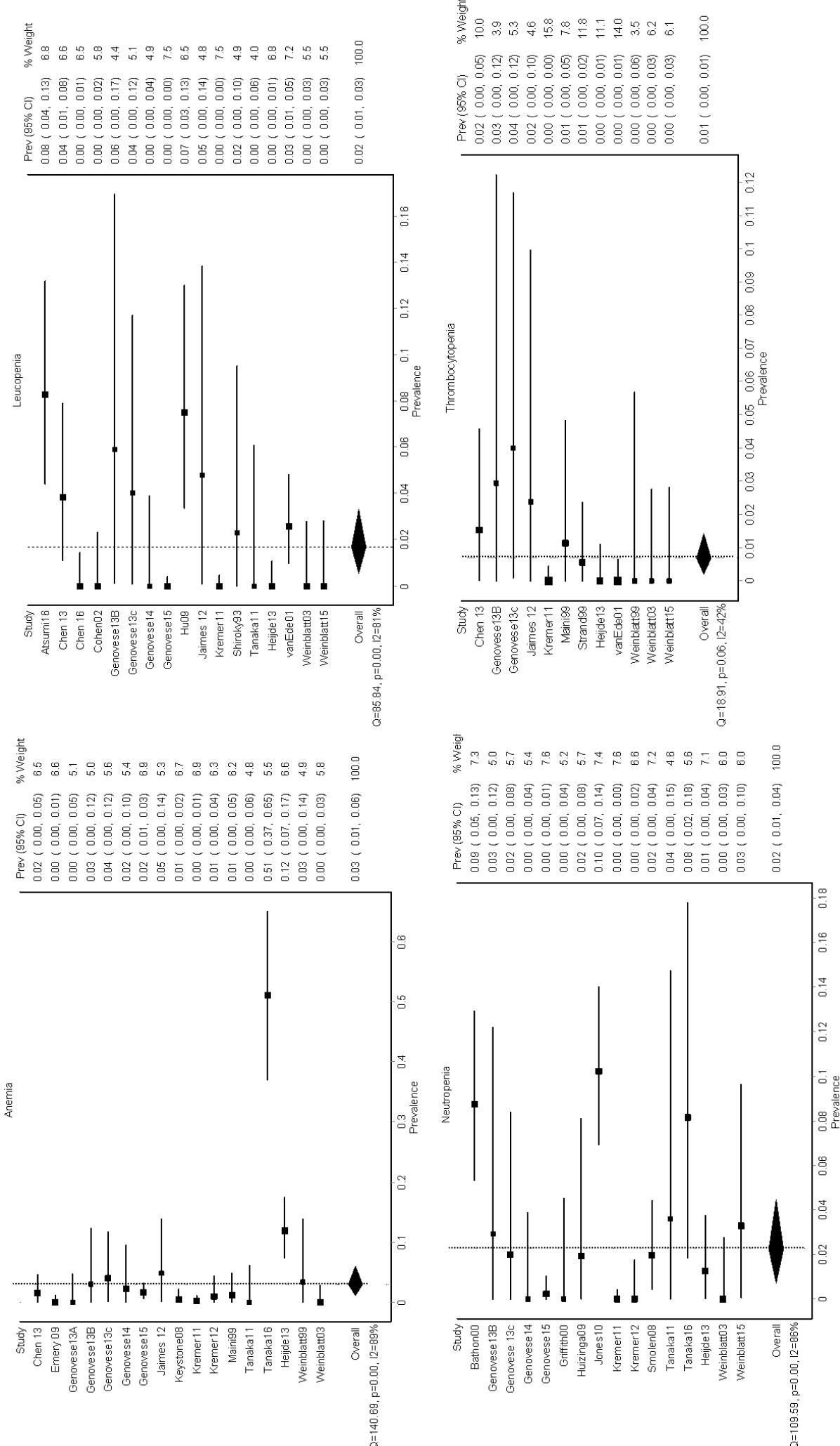
Katheeln Vinny, Zhi Zhang, Danial Solomon

Background/Purpose: Methotrexate (MTX) is known to increase the risk of cytopenias, but the prevalence of hematologic abnormalities among patients taking low dose MTX is poorly defined. We conducted a systematic literature review and meta-analysis to estimate the prevalence of anemia, leukopenia, neutropenia, and thrombocytopenia associated with MTX plus folic acid supplementation among non-oncology patients.

Methods: We searched MEDLINE, PubMed, and Embase from inception to August 2016 for all randomized controlled clinical trials (RCTs) with a MTX monotherapy arm. We excluded RCTs for cancer and included only double-blind studies that reported on hematologic adverse events. Studies were excluded if patients did not receive folic acid or leucovorin supplementation. Most trials used MTX as the comparator arm against newer therapies. Full text articles were assessed by two independent reviewers. Risk of bias was assessed per Cochrane Risk of Bias guidelines including selection, performance, detection, attrition, and reporting bias. Pooled prevalence estimates were calculated using random-effects models. The heterogeneity across studies was tested using Cochran's Q and I^2 .

Results: Of 1601 studies identified, 30 (1.87%) were included that contained data from 3,858 patients with RA; no other rheumatologic conditions were represented. Seventeen trials reported on anemia (N=2,032), 17 reported on leukopenia (N=2,220), 16 reported on neutropenia (N=2,202), and 12 reported on thrombocytopenia (N=1,507). The mean dose of methotrexate was 15.4 (\pm 4.5) mg/week with a maximum dose of 30 mg/week, and 41.9% of subjects were using oral corticosteroids. Trial duration ranged from 12-62 weeks with a mean of 32 (\pm 17) weeks. The pooled prevalence for anemia was 3.05% (95% CI 1.04-5.95%), leukopenia 1.67% (95% CI 0.55-3.31%), neutropenia 2.25% (95% CI 0.74-4.48%), and thrombocytopenia 0.67% (95% CI 0.18-1.42%) (**Figure**). Severe anemia was reported in 4 patients (0.20%), severe neutropenia was reported in 3 patients (0.14%), and no cases of severe leukopenia or thrombocytopenia were reported. The risk of bias assessment showed that most methodological limitations came from a failure to describe randomization procedures (N=23, 76.7%) and selective reporting of only severe hematologic adverse events (N=8, 26.7%). Significant statistical heterogeneity existed across studies for all cytopenias. The I^2 (percentage of variation due to heterogeneity rather than chance) was high: I^2 89% for anemia, 81% for leukopenia, 86% for neutropenia, 42% for thrombocytopenia.

Conclusion: Cytopenias are an uncommon side effect of low-dose MTX with folic acid supplementation among RA patients. Randomized controlled clinical trials vary widely in their reporting of hematologic adverse events, with many failing to report mild and moderate cases. Further research is needed to reach a more precise estimate.



GWAS of gout in patients with hyperuricemia identified many possible new candidate risk alleles

Jing Cui, Zhi Zhang, Elizabeth W. Karlson, Daniel H. Solomon

Harvard Medical School

Division of Rheumatology, Immunology and Allergy, Brigham and Women's Hospital

Background: Virtually all gout patients have high levels of uric acid in the blood (hyperuricemia, HU), but approximately 80% of patients with HU will never develop gout. There are a numbers of GWAS that identified risk alleles for hyperuricemia and/or gout using general population as comparison. This study examines genes that might influence the risk of gout among subjects with HU.

Methods: Subjects with HU (uric acid ≥ 6.5 mg/dl from lab data) and available GWAS data were requested from Partners HealthCare Biobank, a collaborative database that contains linked electronic medical records (EMR), survey data and genomic data for 15,000+ subjects. Gout was defined using a validated EMR algorithm with positive predict value of 0.95. Samples were genotyped from three Illumina MEGA chips. We performed standard quality control (QC) procedures for each file, merged them into one analysis dataset, and applied standard QC again. We restricted our genome-wide association study (GWAS) analysis to the Caucasian population. Principal component analysis (PCA) was performed. Among subjects with HU, association between SNPs and gout was tested using logistic regression assuming a genetic additive model. The first 10 PCAs were utilized as covariates to control for any potential population stratification. We also looked at potential gout and HU SNPs with genome-wide significance among European ancestry from the GWAS Catalog (<https://www.ebi.ac.uk/gwas/>) .

Results: 3146 self-reported Caucasian subjects with HU were identified from the Biobank, with 467 (14.8%) were identified as having gout. The gout group was slightly older (mean \pm SD, 72 ± 11 vs 66 ± 14 , $p < 0.0001$), more likely to be male (79% vs 56%, $p < 0.0001$) compared to non-gout HU group. 793,514 SNPs passed QC and were utilized in the GWAS. The most significant SNP was rs1481012 in *ABCG2* gene with p value of 1.8×10^{-6} , which is a known risk factor for gout. We identified 14 SNPs at $p \sim 10^{-6}$ (see Table); 12 were new SNPs and 2 had been previously identified as gout/HU risk SNPs from the GWAS catalog. 53 SNPs were identified at $p \sim 10^{-5}$; only 1 is a previously identified gout/HU risk SNP. 623 SNPs were identified at $p \sim 10^{-4}$; with only 1 as a known gout/HU SNP.

Conclusion: We carried out a GWAS of gout in patients with HU, and found that some of the risk alleles for gout in the general population are also associated with gout in a population with HU. But, we found many possible new SNPs for gout in HU.

Table Top 20 SNPs associated with gout compared to hyperuricemia

SNP	Chr	BP	Allele	Gene	OR	P
Known gout/HU risk SNPs from previous GWAS with genome-wide significance						
rs1481012	4	89039082	G	ABCG2	0.61	1.8E-06
kgp2887248	4	89052323	T	ABCG2	0.61	1.8E-06
rs4148155	4	89054667	G	ABCG2	0.61	1.8E-06
Not found in previous GWAS (new SNPs identified)						
JHU_2.206047717	2	206047718	A	PARD3B	0.20	1.9E-06
rs7020787	9	12340300	C	-	1.42	1.9E-06
rs73235131	21	23989516	T	-	0.60	2.9E-06
rs2367897	12	72492104	G	TRHDE	0.67	2.9E-06
rs1928873	9	12350564	C	-	0.72	4.4E-06
rs4760822	12	72479611	T	TRHDE	0.67	5.5E-06
JHU_9.12340636	9	12340637	T	-	1.39	6.1E-06
rs7222347	17	72027224	T	-	0.52	6.2E-06
rs7040701	9	12318093	C	-	1.38	8.0E-06
rs143936778	5	34594023	A	-	0.41	8.3E-06
rs113854545	10	53981733	G	PRKG1	0.37	8.3E-06
JHU_12.72507970	12	72507971	T	TPH2	0.68	1.0E-05
rs7721690	5	85185350	A	-	1.68	1.1E-05
rs59996437	3	30845824	T	GADL1	0.20	1.2E-05
JHU_10.102775026	10	102775027	G	PDZD7	0.65	1.2E-05
rs10756369	9	12401666	A	-	0.73	1.5E-05
JHU_12.128023894	12	128023895	G	-	1.38	1.8E-05

Identification of Clinically Relevant Pain Profiles in Individuals with Active RA

Zhi Zhang, Alyssa Wohlfahrt, Bing Lu, Clifton O. Bingham III, Marcy B. Bolster, Wendy Marder, Larry W. Moreland, Kristine Phillips, Tuhina Neogi and Yvonne C. Lee

Background/Purpose: Despite DMARD treatment, many RA patients continue to suffer from pain. Defining distinct pain phenotypes may advance the use of therapies targeted at specific pain mechanisms. This study identified pain profiles among RA patients with active disease, with the goal of informing treatment decisions to improve pain outcomes.

Methods: 146 RA patients with active disease were identified from 5 academic medical centers. Trained assessors performed joint and tender point counts. Patient-reported measures of pain, fatigue, sleep and psychological distress were assessed. Quantitative sensory testing was done to evaluate pain sensitivity (pressure pain thresholds), central pain sensitization (temporal summation) and descending pain inhibition (conditioned pain modulation). A principle components analysis (PCA) was performed to identify variables explaining the most variance. These variables were used in a hierarchical cluster analysis to identify pain phenotypes. General linear and logistic regression models were used to identify differences in clinical characteristics.

Results: Based on PCA, 20 variables were included in the cluster analysis, which identified 3 pain profiles (**Figure**): 1) low pain and temporal summation with low psychological distress, fatigue, and sleep problems (N = 48, 32.9%); 2) moderate pain and high temporal summation with moderate psychological distress, fatigue, and sleep problems (N = 44, 30.1%) and 3) moderate pain and temporal summation with high psychological distress, fatigue and sleep problems (N = 54, 37.0%). Catastrophizing and patient global differed across groups, with the low pain group (cluster 1) having the lowest levels of both (**Table**). NSAID use differed across groups, with the highest frequency of use in the high temporal summation group (cluster 2). Disease duration, CRP and swollen joint count did not differ across groups.

Conclusion: Among RA patients with active disease, 3 pain phenotypes emerged that may inform treatment decisions. Patients with moderate pain and high temporal summation (cluster 2) may benefit from strategies to reduce central sensitization, whereas patients with moderate pain and high psychological distress/fatigue/sleep problems (cluster 3) may benefit from treatments to improve mood and sleep. The finding that NSAID use was highest in the group with high temporal summation is interesting, given a study showing that COX-2 inhibition decreases temporal summation and improves pain (Arendt-Nielsen 2016). Future studies are needed to determine the role of pain phenotypes in directing pain management.

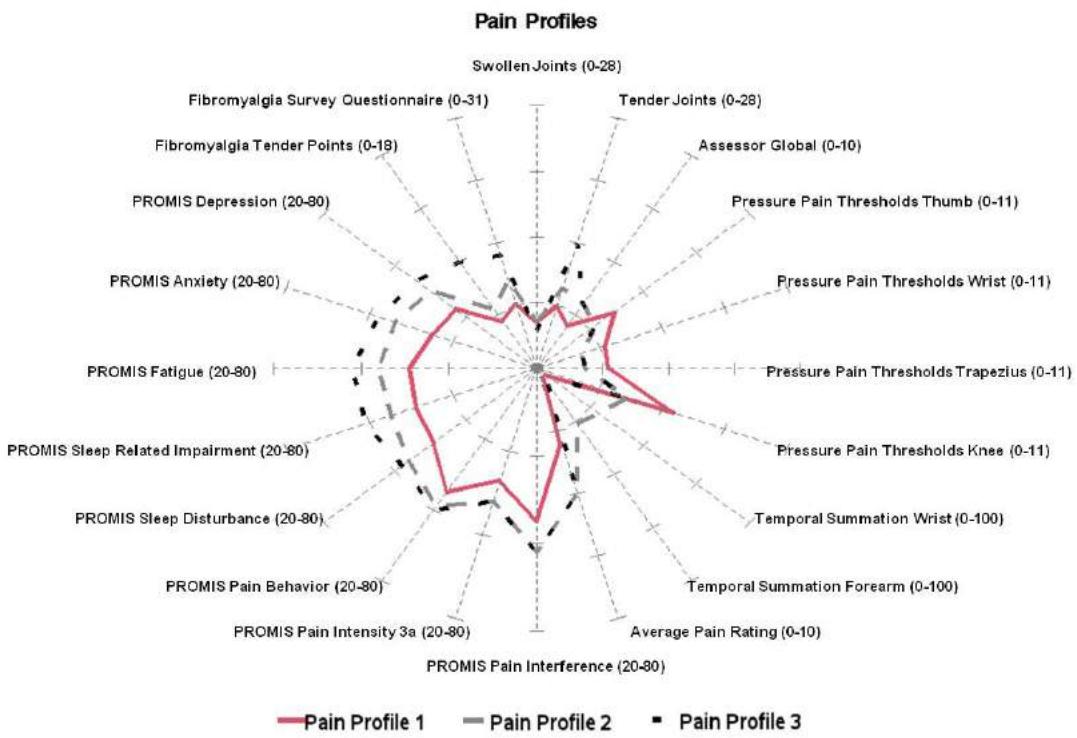


Figure. Radar plot for pain profile 1 (cluster 1, N = 48), pain profile 2 (cluster 2, N = 44), pain profile 3 (cluster 3, N = 54). Each clustering variable is represented by its own spoke with individual axis limits. The shape of the plot represents a map of characteristics for each pain profile.

Table. Clinical characteristics of RA patients with active disease, based on cluster assignment.

	Cluster 1 (N = 48)	Cluster 2 (N = 44)	Cluster 3 (N = 54)	P
Female	72.9%	81.8%	88.9%	0.35
Age, years (SD)	56.7 (12.1)	56.8 (15.9)	51.6 (15.5)	0.13
BMI, mean (SD)	28.4 (6.9)	27.3 (5.2)	29.8 (7.3)	0.18
RA Disease Duration, mean years (SD)	12.2 (14.5)	9.3 (11.7)	10.3 (13.6)	0.56
Seropositive	83.3%	81.8%	72.2%	0.54
CRP (SD)	9.8 (18.5)	9.4 (12.2)	6.6 (8.4)	0.42
Patient Global, 0-10 (SD)	2.9 (2.2)	4.6 (2.2)	4.4 (1.9)	0.0001
Medication Use				
Non-biologic DMARDs	50.0%	45.5%	40.7%	0.64
Biologic DMARDs	27.1%	29.5%	22.2%	0.70
Corticosteroids	33.3%	52.3%	42.6%	0.19
NSAID users	43.8%	63.6%	37.0%	0.03
Pain Catastrophizing Score, 0-52 (SD)	10.1 (8.8)	20.4 (14.0)	22.1 (13.2)	< 0.0001
Conditioned Pain Modulation, (SD)	134.5 (22.6)	133.5 (35.5)	147.6 (44.8)	0.09

Diet and Rheumatoid Arthritis Symptoms: Survey Results From a Rheumatoid Arthritis Registry

SARA K. TEDESCHI, MICHELLE FRITS, JING CUI, ZHI ZACK ZHANG, TAYSIR MAHMOUD, CHRISTINE IANNACCONE, TZU-CHIEH LIN, KAZUKI YOSHIDA, MICHAEL E. WEINBLATT, NANCY A. SHADICK, AND DANIEL H. SOLOMON

Objective. Patients with rheumatoid arthritis (RA) often ask whether specific foods, popularized as inflammatory or antiinflammatory, can improve or worsen their RA. Patients with RA took a survey on diet and RA symptoms, and the survey data were collected and analyzed.

Methods. A dietary survey was mailed to 300 subjects in a single-center RA registry at a large academic center. Subjects were asked about their consumption of 20 foods and whether these foods make their RA symptoms better, worse, or unchanged. Semiannual registry data include demographics, medications, comorbidities, and disease activity scores. Fisher's exact test and Wilcoxon's rank sum tests evaluated associations between subject characteristics from the most recent registry assessment and changes in RA symptoms from specific foods.

Results. Of the 217 subjects (72% response rate), 83% were female; the median RA duration was 17 years (interquartile range 9–27 years), and 58% were taking a biologic disease-modifying antirheumatic drug. Twenty-four percent of subjects reported that foods affect their RA symptoms, with 15% reporting improvement and 19% reporting worsening. Blueberries and spinach were the foods most often reported to improve RA symptoms, while soda with sugar and desserts were those most often reported to worsen RA symptoms. Younger age and noting that sleep, warm room temperature, and vitamin/mineral supplements improve RA were each associated with reporting that foods affect RA symptoms. Medication use, sex, body mass index, smoking, disease duration, disease activity scores, and self-reported RA flares were not associated with reporting that foods affect RA.

Conclusion. Nearly one-quarter of RA subjects with longstanding disease reported that diet had an effect on their RA symptoms.

Introduction

Rheumatoid arthritis (RA) is a chronic autoimmune disease that generally requires the use of disease-modifying antirheumatic drugs (DMARDs) to suppress joint inflammation. However, many RA patients ask whether non-pharmacologic interventions could help to treat their RA. Patient interest in the effect of diet on RA has been noted for decades. In a number of small clinical trials on dietary modifications in RA, modest benefit has been noted

in the use of high-dose omega-3 fatty acids, fasting, a vegetarian diet, and a Mediterranean-type diet (1–3).

In the 21st century, a lay trend of focusing on so-called inflammatory and antiinflammatory foods has emerged. Some patient-oriented medical websites provide examples of such foods, along with advice to consume or avoid these foods for the improvement of RA symptoms (4,5). In the clinical setting, patients often report to their providers that certain foods make their RA better or worse. We hypothesized that patient traits, including RA

The Brigham Rheumatoid Arthritis Sequential Study registry is funded by Amgen, Bristol-Myers Squibb, Crescendo Biosciences, DxTerity, and UCB.

Drs. Tedeschi and Solomon's work was supported by the NIH (grants L30-AR-070514 and K24-AR-055989, respectively).

Sara K. Tedeschi, MD, MPH, Michelle Frits, BA, Jing Cui, MD, PhD, Zhi Zack Zhang, MS, Taysir Mahmoud, BA, Christine Iannaccone, MPH, Tzu-Chieh Lin, PhD, Kazuki Yoshida, MD, MPH, MS, Michael E. Weinblatt, MD, Nancy A. Shadick, MD, MPH, Daniel H. Solomon, MD, MPH; Brigham and Women's Hospital, Boston, Massachusetts.

Drs. Weinblatt and Shadick receive salary support from research grants to their institution from Amgen, Bristol-Myers Squibb, Crescendo Biosciences, DxTerity, and UCB. Dr. Solomon receives salary support from research grants to his institution from Amgen, AstraZeneca, Bristol-Myers Squibb, Eli Lilly, Genentech, and Pfizer.

Address correspondence to Sara K. Tedeschi, MD, MPH, Brigham and Women's Hospital, Division of Rheumatology, Immunology and Allergy, 60 Fenwood Road, Suite 6016, Boston, MA 02115. E-mail: stedeschi1@partners.org.

Submitted for publication October 14, 2016; accepted in revised form February 14, 2017.

Significance & Innovations

- In rheumatoid arthritis (RA) subjects with long-standing RA, 24% noted that specific foods affect their RA symptoms. Self-reported change in RA symptoms did not differ based on medication use or disease activity scores.
- Younger age and reporting that other lifestyle factors (sleep, room temperature, and vitamin/mineral supplements) improve RA symptoms were associated with changes in RA symptoms from foods.
- Of 20 prespecified foods, many of which have been popularized in the lay press as inflammatory or antiinflammatory, blueberries and spinach were most often reported to improve RA symptoms, while soda with sugar and desserts were most often reported to worsen RA symptoms among subjects who reported eating those foods.

disease activity measures and the use of biologic and nonbiologic disease-modifying antirheumatic drugs (DMARDs), differ between patients who note that foods affect their RA symptoms and patients who do not. We surveyed RA patients to learn whether or not specific foods affect their RA symptoms, and tested whether patient traits were associated with a change in RA symptoms from certain foods.

Patients and methods

Study design and cohort. The Brigham RA Sequential Study is a single-center longitudinal RA registry that has enrolled >1,400 subjects since 2003. All subjects were diagnosed with RA by a rheumatologist, and 97% of subjects meet 1987 American College of Rheumatology (ACR) criteria for RA (6). Rheumatoid factor and anti-cyclic citrullinated peptide antibody are tested at enrollment. Patient-reported data collected every 6 months include demographics, RA medications, comorbidities, and self-reported disease activity scores (RA Disease Activity Index and modified Health Assessment Questionnaire). Physician-reported disease activity measures (Disease Activity Score in 28 joints using the C-reactive protein level [DAS28-CRP] and Clinical Disease Activity Index score) are collected annually at a study visit. All elements of this study were approved by the Partners Healthcare Institutional Review Board.

Survey. Between May 2015 and December 2015, a structured dietary survey was mailed to a random sample of 300 subjects who were actively participating in the registry. For each of the 20 foods on the survey, subjects were asked to “please indicate whether it makes your RA unchanged, better, or worse.” Subjects were also given the option of “I do not eat this food.” The 20 foods were included on the survey based on popular beliefs about their

inflammatory properties (milk, cheese, red meat, tomatoes, eggplant, white potatoes, bell or hot peppers, diet soft drinks, and beer) (4,5,7), antiinflammatory properties (fish, spinach, blueberries, strawberries, chocolate, and red wine) (4), or because we hypothesized that some subjects would report worsened RA symptoms after consuming them (soft drinks with sugar, caffeinated coffee, decaffeinated coffee, caffeinated tea, and herbal or decaffeinated tea) (8). Subjects were offered an opportunity to provide free-text entries about additional foods that make their RA better or worse. The survey included a question assessing food avoidance (never, sometimes, or often) due to worsened RA symptoms from food.

Subjects were also asked whether 4 lifestyle/environmental factors improve their RA symptoms. These factors were regular exercise, a good night’s sleep, vitamin/mineral supplements, and warm room temperature. These questions were included as indicators of whether subjects perceive other lifestyle/environmental factors, aside from diet, to affect RA symptoms. Prior work supports beneficial effects of exercise in RA (9) and an association between poor sleep and RA symptoms (10). By contrast, little if any data exist on the impact of vitamin/mineral supplements or warm room temperature on RA symptoms. Subjects were also asked to report medical conditions that could potentially affect their dietary choices (e.g., Crohn’s disease, ulcerative colitis, or acid reflux). Completed surveys were mailed to the registry coordinator and were double-data-entered into REDCap (Research Electronic Data Capture) by 2 research assistants.

Statistical analysis. The primary outcome was reporting that foods affect RA symptoms, which included 3 nonoverlapping groups of subjects: those who reported that foods improve RA, that foods worsen RA, or that some foods improve RA and other foods worsen RA. Responses for each of the 20 specific foods, as well as free-text entries, were taken into account. For each of the 20 foods listed in the survey, we calculated the percentage of subjects reporting either worsened or improved RA symptoms due to that food.

Data on medications and disease activity scores were collected from the registry study visit closest to the date of the dietary survey completion. Many of the continuous variables were non-normally distributed; thus, for consistency, all continuous data were analyzed with nonparametric tests. Descriptive statistics summarized traits among all subjects, and then separately among those who reported that foods affect RA and those who did not. We tested for associations between subject traits and reporting that foods affect RA, using Fisher’s exact test for categorical variables and Wilcoxon’s rank sum test for continuous variables. Statistical significance was defined as a 2-sided *P* value less than 0.05. Analyses were performed using SAS, version 9.4.

Results

Characteristics of the 217 subjects who completed the survey (72.3% response rate) are shown in Table 1. Subjects had longstanding RA, with a median disease duration of 17.0 years (interquartile range [IQR] 9.0–27.0

Table 1. Characteristics of 217 RA subjects within 6 months of completing the dietary survey*

Characteristic	Value
Age, median (IQR), years	65.0 (53.0–71.0)
Female	83.0
White	94.9
Body mass index, median (IQR), kg/m ² †	26.0 (22.2–31.0)
Some college education or greater	84.8
Smoking‡	
Never	56.4
Past	40.6
Current	3.0
Seropositivity (rheumatoid factor and/or anti-CCP)†	72.3
RA duration, median (IQR), years	17.0 (9.0–27.0)
Current RA therapy	
No treatment or NSAIDs only	8.3
Corticosteroids‡	24.9
Nonbiologic DMARD only§	13.4
Triple nonbiologic DMARD¶	0.9
Methotrexate mono or combo therapy	54.8
Biologic DMARD#	58.1
≥1 self-reported flare in past 6 months	51.2
Disease activity scores, median (IQR)†	
DAS28-CRP	2.1 (1.6–3.0)
CDAI	6.0 (2.5–11.5)
M-HAQ	0.4 (0.1–0.8)
RADAI	2.2 (1.0–4.1)
Self-reported medical conditions†	
Coronary artery disease**	5.5
High blood pressure	6.0
High cholesterol	6.0
Type 2 diabetes mellitus	0.5
Crohn's or ulcerative colitis	3.3
Irritable bowel syndrome	5.2
Celiac or gluten sensitivity	3.3
Acid reflux or gastroesophageal reflux disease	33.5
Lactose intolerance	13.1
Food allergy	11.5

* Values are percentages unless otherwise indicated. RA = rheumatoid arthritis; IQR = interquartile range; anti-CCP = anti-cyclic citrullinated peptide; NSAIDs = nonsteroidal antiinflammatory drugs; DMARD = disease-modifying antirheumatic drug; DAS28-CRP = Disease Activity Score in 28 joints using the C-reactive protein level; CDAI = Clinical Disease Activity Index; M-HAQ = modified Health Assessment Questionnaire; RADAI = Rheumatoid Arthritis Disease Activity Index.

† Data missing for body mass index (n = 25), smoking (n = 15), seropositivity (n = 29), DAS28-CRP (n = 21), CDAI (n = 15), M-HAQ (n = 15), RADAI (n = 15), Crohn's or ulcerative colitis (n = 3), irritable bowel syndrome (n = 5), celiac disease or gluten sensitivity (n = 5), acid reflux or gastroesophageal reflux disease (n = 2), lactose intolerance (n = 3), and food allergy (n = 9) categories.

‡ Includes corticosteroid monotherapy and combination therapy with DMARDs.

§ Includes methotrexate, leflunomide, sulfasalazine, and hydroxychloroquine. Category includes subjects taking 1 nonbiologic DMARD only.

¶ Triple therapy: methotrexate plus sulfasalazine plus hydroxychloroquine.

Includes tumor necrosis factor inhibitors, abatacept, rituximab, tocilizumab, and tofacitinib as monotherapy or combination therapy.

** Patient-reported history of myocardial infarction, coronary angioplasty, or coronary artery bypass grafting.

years). Methotrexate was used by 54.8%, either as monotherapy or combination therapy. The use of a biologic DMARD was common, including tumor necrosis factor inhibitors (43.3%), abatacept (8.3%), rituximab (1.4%), tocilizumab (3.2%), and tofacitinib (1.8%). One-fourth of subjects were taking corticosteroids, with a median prednisone dosage of 5 mg daily (IQR 4–10). In 8.3% of subjects, either no RA medications were used or only nonsteroidal antiinflammatory drugs (NSAIDs) were used. Although half of the subjects reported having a flare in the past 6 months, disease activity was low overall, with a median DAS28-CRP score of 2.1 (IQR 1.6–3.0). The frequency of self-reported acid reflux and lactose intolerance was similar to published studies of these conditions as self-reported in the general US population (11,12). The vast majority of subjects reported that exercising regularly (83.7%) and getting a good night's sleep (86.5%) improved RA symptoms. Warm room temperature (35.9%) and vitamin/mineral supplements (34.0%) were less often reported to improve RA symptoms.

Impact of foods on RA symptoms. At least 1 food was reported to affect RA symptoms in 24.0% of subjects. In 10.1% of subjects, some foods improved and some foods worsened RA symptoms. Foods only worsened RA in 9.2% of subjects, and foods only improved RA in 4.6% of subjects. Additionally, 24.3% of all subjects reported avoiding foods (16.2% sometimes and 8.1% often) because they worsen their RA. Subjects who reported eating the specific foods listed on the survey most often noted blueberries (11.1%) and fish (10.9%) as improving RA symptoms (Figure 1). Soda with sugar (12.7%) and desserts (12.4%) were the items most frequently noted to worsen RA symptoms among the subjects who consumed them. Caffeinated coffee and caffeinated tea were each reported to affect RA in less than 5% of subjects.

Associations between subject characteristics and reporting that foods affect RA. Subjects who reported that foods affect RA were younger than those who did not (median age 58.5 versus 66.0 years; *P* = 0.02) (Table 2). Sex, body mass index, smoking status, seropositivity, duration of RA, and the use of corticosteroids, nonbiologic DMARDs, and biologic DMARDs did not differ between those who reported that foods affect RA and those who did not. Clinician-reported disease activity scores, patient-reported disease activity scores, and self-reported flares also did not differ between the 2 groups, nor did self-reported medical conditions that might affect diet (e.g., celiac disease/gluten sensitivity, irritable bowel syndrome). Participants who reported that a good night's sleep (*P* < 0.01), warm room temperature (*P* < 0.01), and vitamin/mineral supplements (*P* < 0.01) improve RA symptoms were more likely to report that foods affect RA.

Discussion

We examined a sample of 217 RA patients in an RA registry to learn how foods affect their RA symptoms. Nearly one-quarter of RA subjects with longstanding

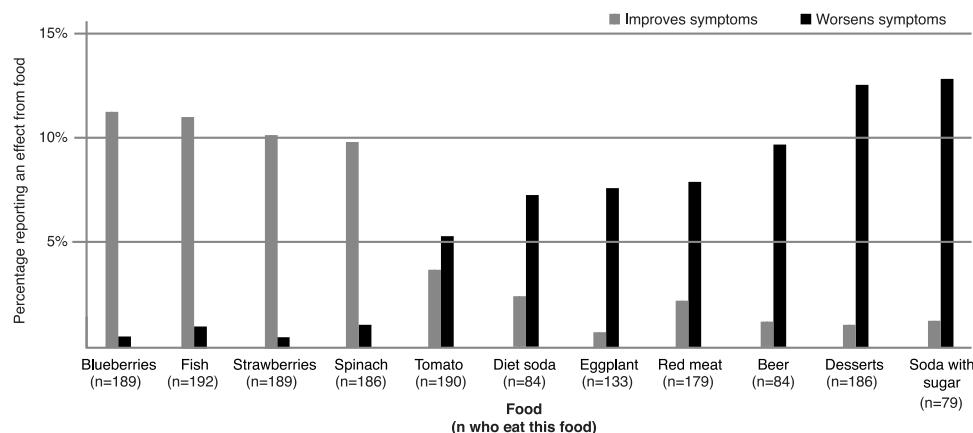


Figure 1. Foods that study subjects reported most often as affecting their rheumatoid arthritis symptoms.

disease reported that foods impacted their RA symptoms, making RA better, worse, or better with some foods and worse with other foods. This group reporting an effect on symptoms was younger (although the median age was still >50 years), and was also more likely to report that other lifestyle/environmental factors improve their RA. Other RA characteristics, including medications, disease duration, and disease activity scores did not differ based on reporting that foods affect RA.

One similar survey on this topic was conducted from 1985–1990 among 704 private practice RA patients (13). In that cohort, specific foods affected RA symptoms in 27.6% of subjects. Foods only worsened RA in 10.7% of respondents, only improved RA in 5.5%, and some foods worsened while other foods improved RA in 9.8%. Similar to the present study, subjects who reported that foods affect RA symptoms were younger, and other traits did not differ between individuals who reported that foods affect RA and those who did not. Our study additionally tested for differences by medication use, self-reported flares, clinician- or patient-reported disease activity measures, and beliefs about other lifestyle factors.

Subgroups of RA patients may differ in whether foods affect RA symptoms. We hypothesized that subjects using biologic DMARDs would be less likely to report that foods affect RA, due to the potent effect of these medications. However, we did not detect a difference in responses based on the use of biologic DMARDs, nor based on use of nonbiologic DMARDs or corticosteroids. The literature is conflicting on whether caffeine leads to decreased efficacy of methotrexate via inhibitory effects on extracellular adenosine (8,14). Of the 54.8% of subjects using methotrexate in our study, none reported that caffeinated coffee or tea made their RA worse (data not shown). It is interesting that, in the era of potent DMARD therapy, a similar percentage of RA subjects report that diet impacts their RA, compared to results from 25 years ago, when NSAIDs, gold, corticosteroids, and hydroxychloroquine were among the mainstays of therapy.

A sizeable percentage of RA patients in this longitudinal registry reported that foods affect their RA symptoms. While many of the foods included in the dietary survey have been popularized as inflammatory or

antiinflammatory, to our knowledge the effect of many of these foods on inflammation has not been formally studied in RA. Our study did not test whether specific foods affect RA disease activity, and we did not find a cross-sectional association between RA disease activity scores and self-reported change in RA symptoms from food. It is possible that certain foods on our survey do have anti-inflammatory effects in RA; for example, a recent randomized trial in juvenile idiopathic arthritis tested the effect of etanercept, etanercept plus daily blueberry juice, and etanercept plus placebo juice on disease activity and circulating interleukin-1 (IL-1) levels (15). At 6 months, the etanercept plus blueberry juice group had reduced IL-1 levels, and a significantly higher percentage of subjects achieved ACR criteria for 20%, 50%, and 70% improvement compared to the other 2 groups. However, we are unable to draw conclusions about the effect of blueberries on RA disease activity in this survey study.

Additionally, we revealed that approximately one-third of subjects feel that vitamin/mineral supplements and warm room temperature improve RA symptoms. These findings have not been investigated or supported by the literature to date. Our study indicates that there is a substantial amount of patient interest in diet and RA symptoms, and highlights the need for prospective dietary studies before RA-specific recommendations can be made.

The present study was cross-sectional, and assessed the self-reported impact of only 20 specific foods. We were thus unable to assess dietary patterns in relation to RA symptoms or disease activity scores. Subjects were participants in a single-center RA registry at a large, academic medical center, were highly educated, and predominantly white; therefore, our findings may not be generalizable to other RA populations. Prospective studies of specific foods and dietary patterns in relation to disease activity scores are needed, especially in the era of potent DMARD therapy.

In conclusion, we found that 24.0% of subjects in a prospective, longitudinal RA registry reported that foods affect their RA symptoms, and 24.3% avoid foods to prevent worsening of their RA. From the list of 20 specific foods, blueberries and spinach were most often noted to improve RA, while soda with sugar and desserts were

Table 2. Comparison of subjects who report that specific foods affect their RA symptoms, and those who do not*

	Yes (n = 52)	No effect (n = 165)	P
Age, median (IQR), years	58.5 (51.0–68.5)	66.0 (54.0–72.0)	0.02
Female	90.4	80.6	0.14
White	96.2	94.6	0.99
BMI, median (IQR), kg/m ² †	25.2 (21.5–31.8)	26.3 (22.3–30.9)	0.55
Some college education or greater	82.7	85.5	0.66
Smoking‡			0.41
Never	62.0	54.6	
Past	38.0	41.5	
Current	0	4.0	
Seropositivity (RF and/or anti-CCP)†	65.1	74.5	0.25
RA duration, median (IQR), years	16.5 (8.5–22.0)	17.0 (9.0–29.0)	0.46
Current RA therapy			
Corticosteroids‡	26.9	24.2	0.72
Nonbiologic DMARD only§	7.7	15.2	0.24
Methotrexate mono- or combination therapy	48.1	57.0	0.27
Biologic DMARD¶	65.4	55.8	0.26
Disease activity scores, median (IQR)†			
DAS28-CRP	2.0 (1.5–3.2)	2.1 (1.6–3.0)	0.68
CDAI	5.5 (2.0–12.0)	6.8 (2.5–11.5)	0.58
M-HAQ	0.5 (0.0–0.8)	0.4 (0.1–0.8)	0.89
RADAi	2.3 (1.0–4.1)	2.1 (0.9–4.1)	0.71
≥1 self-reported flares in past 6 months	61.5	47.9	0.15
Self-reported medical conditions†			
Irritable bowel syndrome	5.9	5.0	0.73
Celiac or gluten sensitivity	7.8	1.9	0.06
Acid reflux or gastroesophageal reflux disease	28.9	35.0	0.50
Lactose intolerance	15.4	12.4	0.64
Food allergy	18.0	9.5	0.13
Other factors reported to improve RA symptomst			
Regular exercise	88.2	82.3	0.39
Good night's sleep	98.1	82.7	< 0.01
Vitamin/mineral supplements	55.8	26.9	< 0.01
Warm room temperature	52.9	30.4	< 0.01

* Values are percentages unless otherwise indicated. P values derived from Fisher's exact test or Wilcoxon's rank sum test. RA = rheumatoid arthritis; IQR = interquartile range; BMI = body mass index; RF = rheumatoid factor; anti-CCP = anti-cyclic citrullinated peptide; DMARD = disease-modifying antirheumatic drug; DAS28-CRP = Disease Activity Score in 28 joints using the C-reactive protein level; CDAI = Clinical Disease Activity Index; M-HAQ = modified Health Assessment Questionnaire; RADAi = Rheumatoid Arthritis Disease Activity Index.

† Data missing for regular exercise (n = 5), good night's sleep (n = 2), vitamins/minerals (n = 5), and room temperature (n = 3). See Table 1 footnote for additional missing-data categories.

‡ Includes corticosteroid monotherapy and combination therapy with DMARDs.

§ Includes methotrexate, leflunomide, sulfasalazine, and hydroxychloroquine. This category includes subjects taking 1 nonbiologic DMARD only.

¶ Includes tumor necrosis factor inhibitors, abatacept, rituximab, tocilizumab, and tofacitinib as monotherapy or combination therapy.

most commonly reported to worsen RA. While we cannot draw strong conclusions based on this survey, past work has shown that greater consumption of sugar-sweetened beverages is associated with increased risk for developing RA (16). A potential link between sugar consumption and inflammation in rheumatoid arthritis warrants further study.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be submitted for publication. Dr.

Tedeschi had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study conception and design. Tedeschi, Weinblatt, Shadick, Solomon.

Acquisition of data. Frits, Mahmoud, Iannaccone.

Analysis and interpretation of data. Tedeschi, Cui, Zhang, Lin, Yoshida.

ROLE OF THE STUDY SPONSOR

Amgen, Bristol-Myers Squibb, Crescendo Biosciences, DxTerity, and UCB provided funding for the Brigham Rheumatoid Arthritis Sequential Study but had no role in the study design or in the collection, analysis, or interpretation of the data, the

writing of the manuscript, or the decision to submit the manuscript for publication. Publication of this article was not contingent upon approval by Amgen, Bristol-Myers Squibb, Crescendo Biosciences, DxTerity, or UCB.

REFERENCES

1. Panush RS, Carter RL, Katz P, Kowsari B, Longley S, Finnie S. Diet therapy for rheumatoid arthritis. *Arthritis Rheum* 1983;26:462–71.
2. Kremer JM, Lawrence DA, Petrillo GF, Litts LL, Mullaly PM, Rynes RI, et al. Effects of high-dose fish oil on rheumatoid arthritis after stopping nonsteroidal antiinflammatory drugs: clinical and immune correlates. *Arthritis Rheum* 1995;38:1107–14.
3. Tedeschi SK, Costenbader KH. Is there a role for diet in the therapy of rheumatoid arthritis? *Curr Rheumatol Rep* 2016;18:23.
4. Arthritis Foundation. Anti-inflammatory diet. URL: <http://www.arthritis.org/living-with-arthritis/arthritis-diet/anti-inflammatory/anti-inflammatory-diet.php>.
5. Arthritis Foundation. Eight food ingredients that can cause inflammation. URL: <http://www.arthritis.org/living-with-arthritis/arthritis-diet/foods-to-avoid-limit/food-ingredients-and-inflammation.php>.
6. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum* 1988;31:315–24.
7. Cleveland Clinic. How an anti-inflammatory diet can relieve pain as you age. URL: <https://health.clevelandclinic.org/2015/11/anti-inflammatory-diet-can-relieve-pain-age/>.
8. Nesher G, Mates M, Zevin S. Effect of caffeine consumption on efficacy of methotrexate in rheumatoid arthritis. *Arthritis Rheum* 2003;48:571–2.
9. Van der Giesen FJ, van Lankveld W, Hopman-Rock M, de Jong Z, Munneke M, Hazes JM, et al. Exploring the public health impact of an intensive exercise program for patients with rheumatoid arthritis: a dissemination and implementation study. *Arthritis Care Res (Hoboken)* 2010;62:865–72.
10. Lee YC, Chibnik LB, Lu B, Wasan AD, Edwards RR, Fosse AH, et al. The relationship between disease activity, sleep, psychiatric distress and pain sensitivity in rheumatoid arthritis: a cross-sectional study. *Arthritis Res Ther* 2009;11:R160.
11. Cohen E, Bolus R, Khanna D, et al. GERD symptoms in the general population: prevalence and severity versus care-seeking patients. *Dig Dis Sci* 2014;59:2488–96.
12. Keith JN, Nicholls J, Reed A, Kafer K, Miller GD. The prevalence of self-reported lactose intolerance and the consumption of dairy foods among African American adults are less than expected. *J Natl Med Assoc* 2011;103:36–45.
13. Tanner SB, Callahan LF, Panush R, Pincus T. Dietary and allergic associations with rheumatoid arthritis: self-report of 704 patients. *Arthritis Rheum* 1990;3:189–95.
14. Benito-Garcia E, Heller JE, Chibnik LB, Maher NE, Matthews HM, Bilics JA, et al. Dietary caffeine intake does not affect methotrexate efficacy in patients with rheumatoid arthritis. *J Rheumatol* 2006;33:1275–81.
15. Zhong Y, Wang Y, Guo J, Chu H, Gao Y, Pang L. Blueberry improves the therapeutic effect of etanercept on patients with juvenile idiopathic arthritis: phase III study. *Tohoku J Exp Med* 2015;237:183–91.
16. Hu Y, Costenbader KH, Gao X, et al. Sugar-sweetened soda consumption and risk of developing rheumatoid arthritis in women. *Am J Clin Nutr* 2014;100:959–67.

Association Between Pain Sensitization and Disease Activity in Patients With Rheumatoid Arthritis: A Cross-Sectional Study

YVONNE C. LEE,¹ CLIFTON O. BINGHAM III,² ROBERT R. EDWARDS,¹ WENDY MARDER,³ KRISTINE PHILLIPS,³ MARCY B. BOLSTER,⁴ DANIEL J. CLAUW,³ LARRY W. MORELAND,⁵ BING LU,¹ ALYSSA WOHLFAHRT,¹ ZHI ZHANG,¹ AND TUHINA NEOGI⁶

Objective. Pain sensitization may contribute to pain severity in rheumatoid arthritis (RA), impacting disease activity assessment. We examined whether pain processing mechanisms were associated with disease activity among RA patients with active disease.

Methods. The study included 139 subjects enrolled in the Central Pain in Rheumatoid Arthritis cohort. Subjects underwent quantitative sensory testing (QST), including assessment of pressure pain thresholds (PPTs) at multiple sites, conditioned pain modulation, and temporal summation. RA disease activity was assessed using the Clinical Disease Activity Index (CDAI) and its components. We examined cross-sectional associations between QST measures and disease activity using linear regression.

Results. Low PPTs (high pain sensitization) at all sites were associated with high CDAI scores ($P \leq 0.03$) and tender joint counts ($P \leq 0.002$). Associations between PPTs and patient global assessments were also seen at most sites. High temporal summation at the forearm (also reflecting high pain sensitization) was significantly associated with high CDAI scores ($P = 0.02$), patient global assessment scores ($P = 0.0006$), evaluator global assessment scores ($P = 0.01$), and tender joint counts ($P = 0.02$). Conversely, conditioned pain modulation (a measure of descending inhibitory pain pathways) was associated only with tender joint count ($P = 0.03$).

Conclusion. High pain sensitization is associated with elevations in disease activity measures. Longitudinal studies are underway to elucidate the cause–effect relationships between pain sensitization and inflammatory disease activity in RA.

INTRODUCTION

Pain is often considered a surrogate marker for inflammatory disease activity in rheumatoid arthritis (RA). It is the single largest determinant of patient assessment of global disease activity (1,2). It is also a prominent component of the American College of Rheumatology (ACR)/European

League Against Rheumatism (EULAR) criteria for remission (3,4). However, pain does not always equal inflammation. Evidence of this was seen in one study, in which the majority of established RA patients with pain (median 3 of 10 in intensity) had a minimum number of swollen joints (5).

Several studies indicate that individuals with RA have abnormalities in peripheral and central nervous system

The contents of this article are solely the responsibility of the authors and do not necessarily represent the official views of the NIH or of Harvard Catalyst or Harvard University and its affiliated academic health care centers.

Supported by the NIH/National Institute of Arthritis and Musculoskeletal and Skin Diseases (grants R01-AR-064850, P60-AR-47785, and R01-AR-062506), Harvard Catalyst/The Harvard Clinical and Translational Science Center (NIH UL1-TR-001102), Harvard University and its affiliated academic health care centers, the Rheumatic Diseases Research Core Center (P30-AR-053503), and the Camille Julia Morgan Arthritis Research and Education Fund.

¹Yvonne C. Lee, MD, MMSc, Robert R. Edwards, PhD, Bing Lu, MD, DrPH, Alyssa Wohlfahrt, BA, Zhi Zhang, MS: Brigham and Women's Hospital, Boston, Massachusetts; ²Clifton O. Bingham III, MD: Johns Hopkins University, Baltimore, Maryland; ³Wendy Marder, MD, Kristine Phillips, MD, Daniel J. Clauw, MD: University of Michigan Medical School, Ann Arbor; ⁴Marcy B. Bolster,

MD: Massachusetts General Hospital, Boston; ⁵Larry W. Moreland, MD: University of Pittsburgh, Pittsburgh, Pennsylvania; ⁶Tuhina Neogi, MD, PhD: Boston University School of Medicine, Boston, Massachusetts.

Dr. Lee has received grant support from Pfizer and owns stock in Express Scripts. Dr. Bolster has received research support from Amgen and Eli Lilly. Dr. Clauw has received consulting fees from Eli Lilly, Nuvo, Cerephex, Tonix, Abbott, Forest Labs, Johnson & Johnson, Merck, Purdue Pharma, Samumed, Zynerba, Astellas Pharma, Williams & Connolly LLP, and Therevance (less than \$10,000 each) and from Pfizer (more than \$10,000), and has received research support from Pfizer, Cypress Biosciences, Forest, Merck, Nuvo, and Cerephex.

Address correspondence to Yvonne C. Lee, MD, MMSc, Division of Rheumatology, Immunology and Allergy, Brigham and Women's Hospital, 60 Fenwood Road, Boston, MA 02115. E-mail: ylee9@partners.org.

Submitted for publication July 20, 2016; accepted in revised form April 18, 2017.

Significance & Innovations

- This is the largest study to comprehensively assess pain sensitization in rheumatoid arthritis (RA), using pressure pain thresholds, temporal summation, and conditioned pain modulation, in a study population that draws from 5 academic medical centers across the US.
- Associations between temporal summation at the forearm (a measure of central sensitization) and Clinical Disease Activity Index score, tender joint count, evaluator global assessment score, and patient global assessment score are reported for the first time.
- Associations between conditioned pain modulation (a measure of descending analgesic pain mechanisms) and measures of disease activity in an RA population are described for the first time.

pain processing, resulting in widespread pain sensitivity. Four cross-sectional studies have examined the relationship between pain thresholds and validated measures of disease activity in RA (6–9). All 4 studies were small ($n \leq 59$), limiting the ability to examine differences in associations between subgroups of patients, such as those with secondary fibromyalgia. Two of the studies included only women, thereby limiting generalizability to men (8,9). In this study, we enrolled 139 patients across 5 sites, including 23 men. In addition, we also assessed conditioned pain modulation as a measure of descending analgesic pain mechanisms. We hypothesized that low pressure pain thresholds (PPTs), low conditioned pain modulation and high temporal summation would be associated with high Clinical Disease Activity Index (CDAI) scores, tender joint counts, and patient global health assessment scores, whereas the association between pain sensitization and measures that emphasize direct assessment of inflammation would be low.

PATIENTS AND METHODS

Study population. The study includes baseline data from the first 139 subjects with complete data on disease activity measures in the Central Pain in Rheumatoid Arthritis (CPIRA) study. CPIRA is a multicenter, prospective, observational study designed to examine the relationship between pain and treatment response in RA. Participants were recruited from 5 US academic medical centers beginning in January 2014. The inclusion criteria were as follows: diagnosis of RA based on the ACR/EULAR 2010 classification criteria (10); starting or switching to a disease-modifying antirheumatic drug (DMARD) due to active RA; and ability to participate in a study visit before taking the first dose of the new DMARD. An exception was made for participants starting methotrexate therapy. These individuals were able to participate after taking 1 dose of medication if they were able to come in for their study visit before taking a second dose. Pharmacodynamic studies indicate that the onset of action of

oral methotrexate for RA is between 3 and 6 weeks, so a single dose of methotrexate should not alter the results of our study (11). For individuals switching to a different DMARD, no washout period was required.

Exclusion criteria were as follows: changing doses of centrally acting pain medications (e.g., amitriptyline, gabapentin, or duloxetine) within 3 months of enrollment; corticosteroid treatment of >10 mg of prednisone or its equivalent; chronic opioid use or any opioid use within 24 hours of testing; diagnosis of a systemic autoimmune disease other than RA; severe Raynaud's phenomenon requiring pharmacologic treatment; severe peripheral vascular disease manifested by claudication or ischemic rest pain; and peripheral neuropathy. All subjects provided written informed consent. The institutional review boards at each site approved the study.

Quantitative sensory testing (QST). All assessors attended a 1-day training session and received in-person instruction on the use of QST. Two of the authors (YCL and RRE) supervised these sessions and ensured that testing measures were standardized across all sites. Site visits were conducted approximately 1 year into the study to ensure that standardized protocols were being followed. Assessments of interrater reliability were performed among a subgroup of assessors ($n = 4$), and the intraclass correlation coefficients (ICCs) ranged from 0.71 to 0.90 for the PPT and temporal summation measures. The ICC for conditioned pain modulation was 0.45. As per Cicchetti (12), ICCs 0.40–0.59 were defined as fair, 0.60–0.74 as good, and 0.75–1.00 as excellent. A comparison of QST measures across sites is shown in Supplementary Table 1 (see Supplementary Table 1, available on the *Arthritis Care & Research* web site at <http://onlinelibrary.wiley.com/doi/10.1002/acr.23266/abstract>).

PPTs. Using a Wagner Force 10 FDX algometer, we obtained PPTs at joint sites (bilateral wrists and knees) and nonjoint sites (bilateral trapezius muscles and thumbnails) in random order, with 3 trials at each site. The 1-cm² rubber algometer probe was placed in the center of each anatomic site by the study staff. The pressure was increased at a rate of 0.50 kgf/second until the stimulus first became painful. The pressure at this point was defined as the PPT. To obtain the mean PPT for each site, we averaged the PPTs obtained on both sides of the body during all 3 trials. Low PPTs at joint sites were considered markers of peripheral sensitization, whereas low PPTs at both joint and nonjoint sites were considered markers of central sensitization (13).

Mechanical temporal summation. Consistent with previous literature (14), temporal summation was assessed using a set of 6 probes, with weighted, flat-end wire tips measuring 0.2 mm in diameter (University of North Carolina, Chapel Hill). The weights ranged 8–256 mN. The probes were tested on the participants by slowly touching the full weight of the probe against the subject's skin at the middle of the wrist (joint site) and then the forearm (nonjoint site). Test taps were performed, beginning with the probe of least weight and sequentially increasing the probe weight until the subject reported a pain rating of 30–40/100 or until the heaviest probe was used. Using this probe, temporal summation was measured by tapping the probe against the skin at the test site 10 times, with each

tap lasting approximately 0.5 seconds and with 1 second between stimuli. The subject was asked to rate his/her pain level on a scale of 0–100 after the first, fifth, and tenth taps. Temporal summation was defined as the difference between the pain level at the tenth tap and the pain level at the first tap for each trial. Three trials were performed at each site. Mean temporal summation measurements at the wrist and the forearm were calculated by averaging the results of the 3 trials. Higher measures of temporal summation were considered to reflect greater central sensitization.

Conditioned pain modulation. Conditioned pain modulation was assessed using a procedure that incorporates a conditioning stimulus (painful stimulus that activates the descending analgesic pain pathways) and a test stimulus (painful stimulus to test the analgesic response to the conditioning stimulus) (15,16). The conditioning stimulus was immersion of the right hand in a cold water bath, maintained between 5°C and 7°C. The test stimulus was pressure applied by an algometer at the left trapezius muscle. An initial PPT was obtained before immersion of the hand in the cold water bath. The subject was then instructed to place his/her hand in the water. After 20 seconds, the PPT at the left trapezius muscle was obtained a second time, immediately before the participant removed his/her hand from the water. If the participant was unable to keep the hand in the water for 20 seconds, the second PPT was measured immediately after the removal of the hand from the water. Conditioned pain modulation was defined as the ratio of the PPT at the second time point over the PPT at the first time point, multiplied by 100 (17). A result of 100 meant that there was no difference between the PPT before the subject was exposed to the conditioning stimulus versus the PPT after the subject was exposed to the conditioning stimulus. Values greater than 100 were indicative of conditioned pain modulation, reflecting increases in PPTs at the second time point compared to PPTs at the first time point. Conversely, lower values were considered to reflect abnormalities in descending pain inhibition.

Assessment of clinical variables. Overall RA disease activity was assessed using the CDAI, a composite measure that includes tender joint count, swollen joint count, patient global assessment, and assessor global assessment (18). We used the CDAI as the primary measure of RA disease activity because serum inflammatory markers (required for the calculation of other validated disease activity measures) will be measured after the full cohort is assembled and are not currently available. Joint counts and assessor global assessments were performed by trained study staff members. For the patient global assessment score, participants were asked to assign a number, using a 100-point numeric rating scale, in response to the question, “Considering all the ways in which your arthritis has affected you, how do you feel your arthritis is today?” Demographic information and RA disease characteristics were obtained using self-report questionnaires. Rheumatoid factor (RF) and anti-cyclic citrullinated peptide (anti-CCP) seropositivity were obtained using a standardized chart review process. Body mass index (BMI) was calculated from height and weight obtained at the time of the study visit. Depression, anxiety, and sleep

disturbance were assessed using Patient-Reported Outcomes Measurement Information System (PROMIS) computerized adaptive tests (19,20). Catastrophizing was assessed using the Pain Catastrophizing Scale (21). Fibromyalgia status was determined according to the ACR 2010 modified preliminary diagnostic criteria, which include meeting a score of ≥ 7 on the Widespread Pain Index and ≥ 5 on the Symptom Severity Scale or 3–6 on the Widespread Pain Index and ≥ 9 on the Symptom Severity Scale (22,23).

Statistical analysis. The primary outcome was RA disease activity, measured by the CDAI. Secondary outcomes included the components of the CDAI, specifically, the tender joint count, swollen joint count, patient global assessment, and assessor global assessment. The main predictors were PPTs at the wrists, knees, trapezius muscles, and thumbnails; temporal summation at the forearm and the wrist; and conditioned pain modulation. Potential confounders included age, sex, BMI, RA disease duration, RF or anti-CCP seropositivity, depression, sleep disturbance, and catastrophizing.

Unadjusted associations between QST measures and clinical disease activity were identified using Pearson correlation coefficients. We examined the association between

Table 1. Participant characteristics (n = 139)*

Characteristic	Value
Age, mean \pm SD years	54.2 \pm 13.6
Female	83.5
Body mass index, mean \pm SD kg/m ²	30.9 \pm 17.3
Seropositive	83.5
Disease duration, mean \pm SD years	9.3 \pm 12.7
CDAI score, mean \pm SD	24.4 \pm 14.0
Tender joint count in 28 joints, mean \pm SD	11.4 \pm 9.2
Swollen joint count in 28 joints, mean \pm SD	5.5 \pm 5.1
Patient global assessment score (0–10), mean \pm SD	5.3 \pm 1.8
Assessor global assessment score (0–10), mean \pm SD	3.7 \pm 2.3
Any DMARD use†	61.2
Biologic DMARD use†	25.2
Synthetic DMARD use†	46.0
NSAID use	48.9
Corticosteroid use	43.2
Pain (0–10 NRS), mean \pm SD	5.1 \pm 2.3
PROMIS depression (T score), mean \pm SD	50.5 \pm 9.1
PROMIS anxiety (T score), mean \pm SD	53.7 \pm 8.7
PROMIS sleep disturbance (T score), mean \pm SD	54.2 \pm 9.2
Pain Catastrophizing Scale, mean \pm SD	18.4 \pm 13.4
Fibromyalgia‡	31.7

* Values are percentages unless otherwise indicated. CDAI = Clinical Disease Activity Index; DMARD = disease-modifying antirheumatic drug; NSAID = nonsteroidal antiinflammatory drug; NRS = numeric rating scale; PROMIS = Patient-Reported Outcomes Measurement Information System.

† Numbers reflect DMARD use within 6 weeks of the baseline assessment, prior to starting their new DMARD.

‡ Defined by the American College of Rheumatology 2010 modified preliminary diagnostic criteria.

QST measures and disease activity using multivariable linear regression, after adjusting for the covariates listed above. The strength of association was assessed using regression coefficients (β). In exploratory analyses, we also examined the role of fibromyalgia as an effect modifier of the relationship between QST measures and disease activity. These analyses were performed using models stratified by fibromyalgia status, as well as models including a QST measure \times fibromyalgia interaction term. The threshold for statistical significance was set as a 2-tailed P value of less than 0.05. We did not adjust for multiple comparisons because this was an observational, hypothesis-screening study, and adjustments for multiple testing limit the ability to identify potentially important findings (24). All analyses were performed using SAS, version 9.4.

RESULTS

Patient characteristics and PPTs. There were 139 RA patients included in the analysis. The mean \pm SD age was 54.2 ± 13.6 years, and 83.5% were women (Table 1). The mean \pm SD baseline CDAI score was 24.4 ± 14.0 , and 31.7% met the ACR 2010 modified preliminary diagnostic criteria for fibromyalgia. The mean \pm SD PPT was lowest at the wrists and trapezius muscles (2.9 ± 1.6) and highest at the knees (5.3 ± 2.7) (Table 2). PPTs were inversely associated with the CDAI, with beta coefficients ranging from -1.29 at the thumbnail to -3.30 at the trapezius ($P \leq 0.03$) (Table 3). PPTs were also significantly associated with the tender joint count, with beta coefficients ranging from -1.09 at the thumbnail to -1.98 at the trapezius ($P \leq 0.002$). PPTs at all sites except the thumbnail were significantly associated with patient global assessment scores ($P \leq 0.04$). In contrast, PPTs were not significantly associated with the swollen joint count.

In stratified analyses, beta coefficients for the association between PPTs and both the CDAI and tender joint counts were generally higher among RA patients who met the 2010 ACR modified preliminary criteria for fibromyalgia (β range = -1.07 , -5.72), compared with those who did not meet the 2010 ACR modified preliminary criteria for fibromyalgia (β range = -0.81 , -3.11). To assess the statistical significance of these differences, we performed exploratory analyses

using multivariable linear regression models including an interaction term for PPT \times fibromyalgia. None of the interaction terms were found to be statistically significant.

Temporal summation. The mean \pm SD values for temporal summation at the wrist and forearm were 15.0 ± 15.3 and 14.0 ± 13.8 , respectively (Table 2). Temporal summation at the forearm was significantly associated with the CDAI score ($\beta = 0.19$; $P = 0.02$), tender joint count ($\beta = 0.11$; $P = 0.02$), patient global assessment score ($\beta = 0.05$; $P = 0.0006$), and assessor global assessment score ($\beta = 0.04$; $P = 0.01$), whereas temporal summation at the wrist was significantly associated only with patient global assessment ($\beta = 0.04$; $P = 0.003$) (Table 4). In analyses stratified by fibromyalgia status, beta coefficients for the association between temporal summation and CDAI score were lower among those with RA and fibromyalgia ($\beta = -0.02$ in forearm and $\beta = -0.01$ in wrist), compared with those with RA alone ($\beta = 0.25$ in forearm and $\beta = 0.23$ in wrist). The interaction terms for temporal summation \times fibromyalgia were not statistically significant.

Conditioned pain modulation. The mean \pm SD conditioned pain modulation ratio was 142.3 ± 39.4 (Table 2). Conditioned pain modulation was associated with tender joint count ($\beta = 0.04$; $P = 0.03$) but not with any other disease activity measure (Table 4). Analyses stratified by fibromyalgia status did not reveal significant differences in the beta coefficients for the associations between conditioned pain modulation and disease activity measures. Interaction terms for conditioned pain modulation \times fibromyalgia were not statistically significant.

DISCUSSION

This study confirms previous findings showing associations between PPTs and composite measures of RA disease activity, tender joint count, evaluator global assessment, and patient global assessment (6,8,25). This study is also the first to report associations between temporal summation at the forearm and CDAI, tender joint count, evaluator global assessment, and patient global assessment. These findings suggest that pain sensitization, reflected by low PPTs and high temporal summation, may contribute to the amplification of patient assessment of disease activity and tender joint count, as well as a perception of higher activity by the evaluator.

To provide clinical context, we compared our results to published data using the same techniques (e.g., same test stimulus and same conditioning stimulus). The median PPT in this population ($2.5\text{--}4.9$ kgf) was lower than that in the general population ($6.2\text{--}9.4$ kgf) and lower than that observed in an RA population with lower disease activity ($5.2\text{--}8.4$ kgf) (15). Temporal summation at the forearm (mean \pm SD 14.0 ± 13.8) was higher in our population compared to healthy controls (mean \pm SD 10.6 ± 11.3) (26). These comparisons should be interpreted with caution, given possible differences in study populations beyond the differences in disease state and disease activity levels.

The observation that low PPTs were associated with high CDAI scores, high tender joint counts, and high patient global

Table 2. Quantitative sensory testing measures (n = 139)*

Measure	Mean \pm SD	Median (IQR)
PPT at wrist, kgf	2.9 ± 1.6	$2.5 (1.9\text{--}3.8)$
PPT at knee, kgf	5.3 ± 2.7	$4.9 (3.0\text{--}7.3)$
PPT at thumbnail, kgf	3.6 ± 1.9	$3.1 (2.4\text{--}4.4)$
PPT at trapezius, kgf	2.9 ± 1.6	$2.5 (1.9\text{--}3.5)$
Temporal summation at wrist†	15.0 ± 15.3	$10.0 (2.7\text{--}23.3)$
Temporal summation at forearm†	14.0 ± 13.8	$11.3 (1.7\text{--}22.0)$
Conditioned pain modulation‡	142.3 ± 39.4	$132.6 (117.7\text{--}155.7)$

* IQR = interquartile range; PPT = pressure pain threshold.

† Calculated as the difference between the maximum pain rating at the tenth tap minus the pain rating at the first tap.

‡ Calculated as PPT₂/PPT₁ \times 100.

Table 3. Relationship of pressure pain thresholds to RA disease activity*

	TJC	SJC	PtGA	EGA	CDAI
Overall study cohort					
Wrist, adj. β	-1.65	-0.46	-0.25	-0.30	-2.66
<i>P</i>	< 0.0001†	0.09	0.04†	0.01†	< 0.0001†
Knee, adj. β	-1.12	-0.17	-0.16	-0.13	-1.58
<i>P</i>	< 0.0001†	0.32	0.03†	0.09	0.0001†
Thumbnail, adj. β	-1.09	0.02	-0.10	-0.11	-1.29
<i>P</i>	0.002†	0.95	0.32	0.31	0.03†
Trapezius, adj. β	-1.98	-0.47	-0.40	-0.44	-3.30
<i>P</i>	< 0.0001†	0.12	0.002†	0.0009†	< 0.0001†
RA patients without FM (n = 95)					
Wrist, adj. β	-1.94	-0.53	-0.33	-0.31	-3.11
<i>P</i>	< 0.0001†	0.09	0.03†	0.03†	< 0.0001†
Knee, adj. β	-0.81	-0.06	-0.16	-0.06	-1.09
<i>P</i>	0.001†	0.73	0.06	0.46	0.009†
Thumbnail, adj. β	-0.99	0.08	-0.13	-0.03	-1.08
<i>P</i>	0.004†	0.76	0.28	0.79	0.07
Trapezius, adj. β	-1.68	-0.30	-0.41	-0.38	-2.78
<i>P</i>	< 0.0001†	0.32	0.005†	0.005†	< 0.0001†
RA patients with FM (n = 44)					
Wrist, adj. β	-1.07	-0.21	-0.08	-0.28	-1.64
<i>P</i>	0.25	0.70	0.73	0.23	0.22
Knee, adj. β	-1.66	-0.06	-0.08	-0.23	-2.02
<i>P</i>	0.009†	0.88	0.62	0.16	0.03†
Thumbnail, adj. β	-1.63	-0.40	-0.04	-0.41	-2.48
<i>P</i>	0.08	0.48	0.86	0.07	0.06
Trapezius, adj. β	-3.79	-0.96	-0.29	-0.68	-5.72
<i>P</i>	0.009†	0.29	0.40	0.07	0.006†

* Adjusted for age, sex, seropositivity, RA disease duration, body mass index, depression, sleep disturbance, and pain catastrophizing. RA = rheumatoid arthritis; TJC = tender joint count; SJC = swollen joint count; PtGA = patient global assessment; EGA = evaluator global assessment; CDAI = Clinical Disease Activity Index; adj. = adjusted; FM = fibromyalgia.

† Statistically significant.

assessment scores, but not with swollen joint counts, is consistent with studies showing that individuals with RA and fibromyalgia score higher on composite disease activity measures and the individual components of tender joint count and patient global assessment (27–30). In our study, beta coefficients for the associations between PPTs and CDAI scores indicated that a 1-unit difference in PPT was associated with a 1.29–3.30 difference in CDAI score. The magnitude of this association was not high, given that the minimum clinically

important difference for the CDAI is 6 for individuals with moderate disease activity and 12 for individuals with high disease activity (31). The strength of this association was higher among individuals with both RA and fibromyalgia, with a beta coefficient of -5.72 for the association between trapezius PPT and CDAI. However, the interaction terms between fibromyalgia and PPTs were not statistically significant when fibromyalgia was examined as a dichotomous variable or as a continuous measure of fibromyalgia symptom

Table 4. Relationship of temporal summation and conditioned pain modulation to RA disease activity in the overall study cohort (n = 139)*

	TJC	SJC	PtGA	EGA	CDAI
Temporal summation (forearm), adj. β	0.11	-0.009	0.05	0.04	0.19
<i>P</i>	0.02†	0.80	0.0006†	0.01†	0.02†
Temporal summation (wrist), adj. β	0.07	-0.02	0.04	0.02	0.11
<i>P</i>	0.10	0.56	0.003†	0.10	0.12
Conditioned pain modulation, adj. β	0.04	-0.001	-0.001	-0.004	0.03
<i>P</i>	0.03†	0.91	0.77	0.47	0.27

* Adjusted for age, sex, seropositivity, RA disease duration, body mass index, depression, sleep disturbance, and pain catastrophizing. RA = rheumatoid arthritis; TJC = tender joint count; SJC = swollen joint count; PtGA = patient global assessment; EGA = evaluator global assessment; CDAI = Clinical Disease Activity Index; adj. = adjusted.

† Statistically significant.

severity. The lack of statistical significance may reflect limited statistical power, given the small number of individuals with fibromyalgia.

High temporal summation at the forearm was significantly associated with high CDAI scores, tender joint counts, evaluator global assessment scores, and patient global assessment scores, but temporal summation at the wrist was associated only with patient global assessment. The beta coefficients for the association between temporal summation and CDAI score ranged from 0.07 at the wrist to 0.11 at the forearm, indicating that a 1-unit difference in temporal summation was associated with an increase in CDAI score of 0.07–0.11. Thus, a large difference in temporal summation is needed to see a relatively small difference in CDAI score.

It was surprising that temporal summation at the wrist was not associated with disease activity measures, because the wrist is a site commonly affected by inflammation in RA. Thus, if anything, we expected stronger associations between temporal summation at the wrist and disease activity measures. One explanation could be that our measure of temporal summation was not sufficiently sensitive. Many subjects did not find the punctate probes to be painful, and the distribution of temporal summation, both at the forearm and the wrist, was right-skewed. We were not able to use a higher-weight probe due to skin fragility in a number of subjects. Compared to other study populations, in this RA population skin fragility may be a larger problem due to chronic corticosteroid use.

To our knowledge, only one other study has examined the association between temporal summation and disease activity measures in RA. Using a temporal summation protocol involving cuff pressure algometry, Vladimirova et al assessed temporal summation at the leg in 38 women with active RA and found no association between the temporal summation index and tender joint count, swollen joint count, or Disease Activity Score in 28 joints (9). A study of 1,111 individuals in the Multicenter Osteoarthritis Study, however, found differences in associations between temporal summation of mechanical stimuli at affected versus unaffected body sites and magnetic resonance imaging-based evaluation of inflammation (32). Over 24 months, this study noted a stronger association between knee effusions and incident temporal summation at the affected site than between knee effusions and incident temporal summation at an unaffected site. Additional studies, using a different method of temporal summation, may be helpful in further elucidating the association between temporal summation and disease activity measures in RA.

Contrary to the associations observed between PPTs and disease activity measures and temporal summation and disease activity measures, conditioned pain modulation was associated with tender joint count and not with any other disease activity measure. The lack of association may be due to several factors, including statistical chance (false negative) and/or technical issues in the assessment of conditioned pain modulation. The ICC for conditioned pain modulation was 0.45, which was lower than the ICCs for the other QST measures, indicating a lower level of reproducibility compared to the other QST measures. In addition, the magnitude of conditioned pain modulation may have been affected by the choice of test and conditioning

stimuli. For example, using cold as the test stimulus may be more sensitive than pressure, given that the cold pressor task was also used as the conditioning stimulus. However, a meta-analysis of conditioned modulation paradigms in populations with chronic pain did not find that the type of test or conditioning stimulus type significantly influenced the effect size (33). Additional studies, using different conditioned pain modulation paradigms, are needed to replicate this finding.

Another possibility for the lack of association between conditioned pain modulation and disease activity measures is that conditioned pain modulation reflects a different type of pain pathway (34). While temporal summation is thought to reflect the facilitation of ascending nociceptive processing, conditioned pain modulation is considered a measure of the descending inhibitory pain pathways (35). We expected impaired conditioned pain modulation to be associated with heightened measures of disease activity due to enhanced pain sensitivity. However, among individuals with high inflammatory disease activity, peripheral inflammation may serve as an endogenous conditioning stimulus that activates the descending analgesic pain mechanisms. Thus, in some individuals, impairments in conditioned pain modulation may be associated with elevations in disease activity measures, while, in others, heightened conditioned pain modulation may be associated with decreases in disease activity measures. Longitudinal assessment of conditioned pain modulation before and after the onset of inflammation would be useful in disentangling these relationships. We are continuing to follow the individuals in this study longitudinally, as they are started on new DMARDs, which provides an opportunity to identify changes in conditioned pain modulation with improvements in inflammation.

Strengths of this study include the comprehensive assessment of PPTs, temporal summation, and conditioned pain modulation. To our knowledge, this is the largest study of QST in RA, and the only study to assess PPTs, temporal summation, and conditioned pain modulation, while also characterizing inflammatory disease activity and psychosocial factors. An additional strength is the assessment of secondary fibromyalgia in this RA cohort. However, misclassification may exist since the ACR 2010 modified preliminary diagnostic criteria for fibromyalgia are based on self-reported pain in 19 areas, fatigue, nonrestorative sleep, and cognitive symptoms (22). Although the 19 areas are nonjoint sites, RA patients may find it difficult to distinguish between pain at different locations.

Limitations of this study include the cross-sectional design, which precludes conclusions involving the directionality of associations between QST measures and disease activity. Longitudinal data collection is ongoing, and analyses to examine associations between baseline QST measures and changes in inflammatory serum markers and composite RA disease activity measures in response to DMARD therapy are planned. The heterogeneity in the assessment of QST measures across sites may be another limitation. These assessments can be sensitive to variations in study procedures. We have made efforts to standardize protocols, including an intensive training session before the start of the study and visiting each site approximately 1 year into the study to ensure that there was no drift in technique.

ICCs between the master study assessor and 3 other study assessors were in the fair to excellent range. When comparing QST measurements across sites, PPTs at the knee and trapezius were the only measures that differed significantly across sites (see Supplementary Table 1, available on the *Arthritis Care & Research* web site at <http://onlinelibrary.wiley.com/doi/10.1002/acr.23266/abstract>). This may reflect variations in testing procedures across sites. Alternatively, this difference may reflect differences in study populations across sites. To address these concerns, we included study site as a covariate in all analyses.

Although we included many potential confounders of the relationship between QST measures and inflammatory disease activity in our models, the potential for residual confounding remains. While we performed a large number of statistical analyses, we avoided adjustment for multiple comparisons in accordance with what has been advocated in epidemiologic research (24). We made a conscious effort to highlight only the associations that were consistent across the majority of body sites or disease activity measures.

In conclusion, pain sensitization, demonstrated by low PPTs and high temporal summation values at the forearm, were associated with high CDAI scores. These findings highlight the importance of understanding pain sensitization in RA, particularly as it relates to inflammatory disease assessment. Additional studies are needed to better understand the clinical impact of pain sensitization on the efficacy of RA treatment.

ACKNOWLEDGMENTS

The authors thank the study coordinators at all of the sites, including Cassandra Corrigan, Agnes Zak, and Dee Luo (Brigham and Women's Hospital); Malini Moni and Grazyna Purwin (Johns Hopkins University); Alieysa Patel and Melanie Woods (University of Michigan); Joyce Goggins (Boston University); and Laurie Hope and Kelly Reckley (University of Pittsburgh). The authors also acknowledge all of the referring physician and patient participants.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be submitted for publication. Dr. Lee had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study conception and design. Lee, Bingham III, Edwards, Phillips, Bolster, Clauw, Moreland, Neogi.

Acquisition of data. Marder, Wohlfahrt.

Analysis and interpretation of data. Lee, Bingham III, Edwards, Marder, Phillips, Bolster, Clauw, Moreland, Lu, Zhang, Neogi.

REFERENCES

1. Furu M, Hashimoto M, Ito H, Fujii T, Terao C, Yamakawa N, et al. Discordance and accordance between patient's and physician's assessments in rheumatoid arthritis. *Scand J Rheumatol* 2014;43:291–5.
2. Studenic P, Radner H, Smolen JS, Aletaha D. Discrepancies between patients and physicians in the perception of rheumatoid arthritis disease activity. *Arthritis Rheum* 2012;64:2814–23.
3. Studenic P, Smolen JS, Aletaha D. Near misses of ACR/EULAR criteria for remission: effects of patient global assessment in Boolean and index-based definitions. *Ann Rheum Dis* 2012;71:1702–5.
4. Felson DT, Smolen JS, Wells G, Zhang B, van Tuyl LH, Funovits J, et al. American College of Rheumatology/European League Against Rheumatism provisional definition of remission in rheumatoid arthritis for clinical trials. *Arthritis Rheum* 2011;63:573–86.
5. Lee YC, Frits ML, Iannaccone CK, Weinblatt ME, Shadick NA, Williams DA, et al. Subgrouping of patients with rheumatoid arthritis based on pain, fatigue, inflammation, and psychosocial factors. *Arthritis Rheumatol* 2014;66:2006–14.
6. Joharatnam N, McWilliams DF, Wilson D, Wheeler M, Pande I, Walsh DA. A cross sectional study of pain sensitivity, disease activity assessment, mental health and fibromyalgia status in rheumatoid arthritis. *Arthritis Res Ther* 2015;17:11.
7. Konttinen YT, Honkanen VE, Gronblad M, Keinonen M, Santavirta N, Santavirta S. The relation of extraarticular tenderness to inflammatory joint disease and personality in patients with rheumatoid arthritis. *J Rheumatol* 1992;19:835–5.
8. Lee YC, Chibnik LB, Lu B, Wasan AD, Edwards RR, Fossel AH, et al. The relationship between disease activity, sleep, psychiatric distress and pain sensitivity in rheumatoid arthritis: a cross-sectional study. *Arthritis Res Ther* 2009;11:R160.
9. Vladimirova N, Jespersen A, Bartels EM, Christensen AW, Bliddal H, Danneskiold-Samsoe B. Pain sensitisation in women with active rheumatoid arthritis: a comparative cross-sectional study. *Arthritis* 2015;2015:434109.
10. Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, Bingham CO III, et al. 2010 rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. *Arthritis Rheum* 2010;62:2569–81.
11. Williams HJ, Willkens RF, Samuelson CO Jr, Alarcón GS, Guttaduria M, Yarboro C, et al. Comparison of low-dose oral pulse methotrexate and placebo in the treatment of rheumatoid arthritis: a controlled clinical trial. *Arthritis Rheum* 1985;28:721–30.
12. Cicchetti DV. Guidelines, criteria and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol Assess* 1994;6:284–90.
13. Graven-Nielsen T, Arendt-Nielsen L. Assessment of mechanisms in localized and widespread musculoskeletal pain. *Nat Rev Rheumatol* 2010;6:599–606.
14. Rolke R, Magerl W, Campbell KA, Schalber C, Caspary S, Birklein F, et al. Quantitative sensory testing: a comprehensive protocol for clinical trials. *Eur J Pain* 2006;10:77–88.
15. Granot M, Weissman-Fogel I, Crispel Y, Pud D, Granovsky Y, Sprecher E, et al. Determinants of endogenous analgesia magnitude in a diffuse noxious inhibitory control (DNIC) paradigm: do conditioning stimulus painfulness, gender and personality variables matter? *Pain* 2008;136:142–9.
16. Lee YC, Lu B, Edwards RR, Wasan AD, Nassikas NJ, Clauw DJ, et al. The role of sleep problems in central pain processing in rheumatoid arthritis. *Arthritis Rheum* 2013;65:59–68.
17. Edwards RR, Mensing G, Cahalan C, Greenbaum S, Narang S, Belfer I, et al. Alteration in pain modulation in women with persistent pain after lumpectomy: influence of catastrophizing. *J Pain Symptom Manage* 2013;46:30–42.
18. Aletaha D, Nell VP, Stamm T, Uffmann M, Pflugbeil S, Machold K, et al. Acute phase reactants add little to composite disease activity indices for rheumatoid arthritis: validation of a clinical activity score. *Arthritis Res Ther* 2005;7:R796–806.
19. Pilkonis PA, Yu L, Dodds NE, Johnston KL, Maihoefer CC, Lawrence SM. Validation of the depression item bank from the Patient-Reported Outcomes Measurement Information System (PROMIS) in a three-month observational study. *J Psychiatr Res* 2014;56:112–9.
20. Buysse DJ, Yu L, Moul DE, Germain A, Stover A, Dodds NE, et al. Development and validation of patient-reported outcome

- measures for sleep disturbance and sleep-related impairments. *Sleep* 2010;33:781–92.
- 21. Sullivan MJ, Bishop SR, Pivik J. The Pain Catastrophizing Scale: development and validation. *Psychol Assess* 1995;7:524–32.
 - 22. Wolfe F, Clauw DJ, Fitzcharles MA, Goldenberg DL, Hauser W, Katz RS, et al. Fibromyalgia criteria and severity scales for clinical and epidemiological studies: a modification of the ACR preliminary diagnostic criteria for fibromyalgia. *J Rheumatol* 2011;38:1113–22.
 - 23. Wolfe F, Smythe HA, Yunus MB, Bennett RM, Bombardier C, Goldenberg DL, et al. The American College of Rheumatology 1990 criteria for the classification of fibromyalgia: report of the Multicenter Criteria Committee. *Arthritis Rheum* 1990;33:160–72.
 - 24. Rothman KJ. No adjustments are needed for multiple comparisons. *Epidemiology* 1990;1:43–6.
 - 25. Pollard LC, Ibrahim F, Choy EH, Scott DL. Pain thresholds in rheumatoid arthritis: the effect of tender point counts and disease duration. *J Rheumatol* 2012;39:28–31.
 - 26. Campbell CM, Carroll CP, Kiley K, Han D, Haywood C Jr., Lanzkron S, et al. Quantitative sensory testing and pain-evoked cytokine reactivity: comparison of patients with sickle cell disease to healthy matched controls. *Pain* 2016;157:949–56.
 - 27. Wolfe F, Cathey MA, Kleinheksel SM. Fibrositis (fibromyalgia) in rheumatoid arthritis. *J Rheumatol* 1984;11:814–8.
 - 28. Coury F, Rossat A, Tebib A, Letroublon MC, Gagnard A, Fantino B, et al. Rheumatoid arthritis and fibromyalgia: a frequent unrelated association complicating disease management. *J Rheumatol* 2009;36:58–62.
 - 29. Duran J, Combe B, Niu J, Rincheval N, Gaujoux-Viala C, Felson DT. The effect on treatment response of fibromyalgic symptoms in early rheumatoid arthritis patients: results from the ESPOIR cohort. *Rheumatology (Oxford)* 2015;54:2166–70.
 - 30. Lee YC, Hackett J, Frits M, Iannaccone CK, Shadick NA, Weinblatt ME, et al. Multibiomarker disease activity score and C-reactive protein in a cross-sectional observational study of patients with rheumatoid arthritis with and without concomitant fibromyalgia. *Rheumatology (Oxford)* 2016;55:640–8.
 - 31. Curtis JR, Yang S, Chen L, Pope JE, Keystone EC, Haraoui B, et al. Determining the minimally important difference in the Clinical Disease Activity Index for improvement and worsening in early rheumatoid arthritis patients. *Arthritis Care Res (Hoboken)* 2015;67:1345–53.
 - 32. Neogi T, Guermazi A, Roemer F, Nevitt MC, Scholz J, Arendt-Nielsen L, et al. Association of joint inflammation with pain sensitization in knee osteoarthritis: the Multicenter Osteoarthritis Study. *Arthritis Rheumatol* 2016;68:654–61.
 - 33. Lewis GN, Heales L, Rice DA, Rome K, McNair PJ. Reliability of the conditioned pain modulation paradigm to assess endogenous inhibitory pain pathways. *Pain Res Manag* 2012;17:98–102.
 - 34. Yarnitsky D. Role of endogenous pain modulation in chronic pain mechanisms and treatment. *Pain* 2015;156 Suppl 1:S24–31.
 - 35. Suzan E, Midbari A, Treister R, Haddad M, Pud D, Eisenberg E. Oxycodone alters temporal summation but not conditioned pain modulation: preclinical findings and possible relations to mechanisms of opioid analgesia. *Pain* 2013;154:1413–8.