Universal Convexification via Risk-Aversion

Krishnamurthy Dvijotham
Dept of Computer Science & Engg
University of Washington
Seattle, WA 98195

Maryam Fazel
Dept of Electrical Engg
University of Washington
Seattle, WA 98195

Emanuel Todorov
Dept of Computer Science & Engg
Dept of Applied Mathematics
University of Washington
Seattle, WA 98195

Abstract

We develop a framework for convexifying a general class of optimization problems. We analyze the suboptimality of the solution to the convexified problem relative to the original nonconvex problem, and prove additive approximation guarantees under some assumptions. In simple settings, the convexification procedure can be applied directly and standard optimization methods can be used. In the general case we rely on stochastic gradient algorithms, whose convergence rate can be bounded using the convexity of the underlying optimization problem. We then extend the framework to a general class of discrete-time dynamical systems where our convexification approach falls under the paradigm of risk-sensitive Markov Decision Processes. We derive the first model-based and model-free policy gradient optimization algorithms with guaranteed convergence to the optimal solution. We also present numerical results in different machine learning applications.

1 INTRODUCTION

It has been said that the watershed in optimization is not between linearity and nonlinearity, but between convexity and nonconvexity. In this paper we develop a framework for convexifying a general class of optimization problems (section 3), turning them into problems that can be solved with efficient convergence guarantees. The convexification approach may change the problem drastically in some cases, which is not surprising since most nonconvex optimization problems are NP-hard and cannot be reduced to solving convex optimization problems. However, under additional assumptions, we obtain guarantees that bound the suboptimality of the solution relative to the original non-convex problem (section 3.2). We adapt stochastic gradient methods to solve the resulting problems (section 4) and prove convergence guarantees that bound the distance to optimality as a function of the number of iterations. In section 5, we extend the framework to arbitrary dynamical systems and derive the first policy optimization approach with guaranteed convergence to the global optimum. The control problems we study fall under the framework of risk-sensitive control. The condition required for convexity is a natural one, relating the control cost, risk factor and noise covariance. It is very similar to the condition that reduces stochastic optimal control to a linear problem [Fleming and Mitter, 1982, Kappen, 2005, Todorov, 2009], which in turn has given rise to path-integral and other specialized methods for control [Theodorou et al., 2010a, Broek et al., 2010, Dvijotham and Todorov, 2011, Toussaint, 2009].

Smoothing with noise is a relatively common approach to simplify difficult optimization problems. It has been used in computer vision [Rangarajan, 1990] heuristically and formalized in recent work [Mobahi, 2012] where it is shown that under certain assumptions, adding sufficient noise eventually leads to a convex optimization problem. In contrast, we show that for any noise level one can choose the degree of risk-aversion so as to obtain a convex problem. This procedure may destroy minima that are not robust to perturbations – which is not necessarily undesirable, because in many applications it makes sense to seek a robust solution rather than the best one.

2 NOTATION

Gaussian random variables are denoted by $\omega \sim \mathcal{N}(\mu, \Sigma)$, where μ is the mean and Σ the covariance matrix. Given a random variable $\omega \in \Omega$ with distribution P and a function $h : \Omega \mapsto \mathbf{R}$, the expected value of the random variable $h(\omega)$ is written as $\underset{\omega \sim P}{\mathrm{E}}[h(\omega)]$. Whenever it is clear from the context, we will drop

the subscript on the expected value and simply write $\mathrm{E}\left[h\left(\omega\right)\right]$. We differ from convention here slightly by denoting random variables with lowercase letters and reserve uppercase letters for matrices. We will frequently work with Gaussian perturbations of a function, so we denote:

$$\bar{h}\left(\theta\right) = \operatorname*{E}_{\omega\sim\mathcal{N}(0,\Sigma)}\left[h\left(\theta+\omega\right)\right], \tilde{h}_\omega\left(\theta\right) = h\left(\theta+\omega\right)-\bar{h}\left(\theta\right).$$

I_n denotes the $n \times n$ identity matrix. Unless stated otherwise, $\|\cdot\|$ refers to the Euclidean ℓ_2 norm. Given a convex set $\mathcal{C}$, we denote the projection onto $\mathcal{C}$ of x by $\mathrm{Proj}_{\mathcal{C}}\left(x\right)$: $\mathrm{Proj}_{\mathcal{C}}\left(x\right) = \mathsf{argmin}_{y\in\mathcal{C}}\|y-x\|$. The maximum eigenvalue of a symmetric matrix M is denoted by $\lambda_{\max}\left(M\right)$. Given symmetric matrices A, B, the notation $A \succeq B$ means that $A - B$ is a positive semidefinite matrix. Given a positive definite matrix $M \succ 0$, we denote the metric induced by M as $\|x\|_M = \sqrt{x^T M x}$.

3   GENERAL OPTIMIZATION PROBLEMS

We study optimization problems of the form:

$$\min_{\theta\in\mathcal{C}} g\left(\theta\right) \tag{1}$$

where g is an arbitrary function and $\mathcal{C} \subset \mathbf{R}^k$ is a convex set. We do not assume that g is convex so the above problem could be a nonconvex optimization problem. In this work, we convexify this problem by decomposing $g\left(\theta\right)$ as follows: $g(\theta) = f(\theta) + \frac{1}{2}\theta^T R\theta$ and perturbing f with Gaussian noise. Optimization problems of this form are very common in machine learning (where R corresponds to a regularizer) and control (where R corresponds to a control cost). The convexified optimization problem is:

$$\min_{\theta\in\mathcal{C}} f_\alpha\left(\theta\right) + \frac{1}{2}\theta^T R\theta \tag{2}$$

where $R \succeq 0$ and

$$f_\alpha\left(\theta\right) = \frac{1}{\alpha}\log\left(\operatorname*{E}_{\omega\sim\mathcal{N}(0,\Sigma)}\left[\exp\left(\alpha f\left(\theta+\omega\right)\right)\right]\right).$$

This kind of objective is common in risk-averse optimization. To a first order Taylor expansion in α, the above objective is equal to $\mathrm{E}\left[f\left(\theta+\omega\right)\right] + \alpha\,\mathrm{Var}\left(f\left(\theta+\omega\right)\right)$, indicating that increasing α will make the solution more robust to Gaussian perturbations. α is called the risk-factor and is a measure of the risk-aversion of the decision maker. Larger values of α will reject solutions that are not robust to Gaussian perturbations.

In order that the expectation exists, we require that f is bounded above:

$$\text{For all } \theta \in \mathbf{R}^k, \ f\left(\theta\right) \leq M < \infty. \tag{3}$$

We implicitly make this assumption throughout this paper in all the stated results. Note that this is not a very restrictive assumption, since, given any function g with a finite minimum, one can define a new objective $g' = \min\left(g, \bar{m}\right)$, where $\bar{m}$ is an upper bound on the minimum (say the value of the function at some point), without changing the minimum. Since the convex quadratic is non-negative, f is also bounded above by $\bar{m}$ and hence $0 < \exp\left(\alpha f\left(\theta+\omega\right)\right) \leq \exp\left(\alpha\bar{m}\right)$. This ensures that $f_\alpha\left(\theta\right)$ is finite. Some results will require differentiability, and we can preserve this by defining g' using a soft-min: For example, $g'\left(x\right) = \bar{m}\tanh\left(\frac{g(x)}{\bar{m}}\right)$.

Theorem 3.1 (Universal Convexification). *The optimization problem* (2) *is a convex optimization problem whenever $\mathcal{C}$ is a convex set and $\alpha R \succeq \Sigma^{-1}$.*

Proof. Writing out the objective function in (2) and scaling by α, we get:

$$\log\left(\operatorname*{E}_{\omega\sim\mathcal{N}(0,\Sigma)}\left[\exp\left(\alpha f(\theta+\omega)\right)\right]\right) + \frac{\alpha}{2}\theta^T R\theta \tag{4}$$

Writing out the expectation, we have:

$$\operatorname*{E}_{\omega\sim\mathcal{N}(0,\Sigma)}\left[\exp\left(\alpha f(\theta+\omega)\right)\right] = \operatorname*{E}_{y\sim\mathcal{N}(\theta,\Sigma)}\left[\exp\left(\alpha f(y)\right)\right]$$

$$\propto \int\exp\left(-\frac{1}{2}(y-\theta)^T\Sigma^{-1}(y-\theta)\right)\exp\left(\alpha f(y)\right)\mathrm{d}y$$

where we omitted the normalizing constant $\left(\sqrt{(2\pi)^n\det\left(\Sigma\right)}\right)^{-1}$ in the last step. Thus, exponentiating (4), we get (omitting the normalizing constant):

$$\int\exp\left(\frac{\alpha}{2}\theta^T R\theta - \frac{1}{2}(y-\theta)^T\Sigma^{-1}(y-\theta) + \alpha f(y)\right)\mathrm{d}y$$

The term inside the exponent can be rewritten as

$$\frac{1}{2}\theta^T\left(\alpha R - \Sigma^{-1}\right)\theta + \theta^T\Sigma^{-1}y + \alpha f\left(y\right) - \frac{1}{2}y^T\Sigma^{-1}y$$

Since $\alpha R \succeq \Sigma^{-1}$, this is a convex quadratic function of θ for each y. Thus, the overall objective is the composition of a convex and increasing function $\log\mathrm{E}\left[\exp\left(\cdot\right)\right]$ and a convex quadratic and is hence convex [Boyd and Vandenberghe, 2004]. Since $\mathcal{C}$ is convex, the overall problem is a convex optimization problem. □

3.1   INTERPRETATION

Theorem 3.1 is a surprising result, since the condition for convexity does not depend in any way on the

properties of f (except for the boundedness assumption (3)), but only on the relationship between the quadratic objective R, the risk factor α and the noise level Σ. In this section, we give some intuition behind the result and describe why it is plausible that this is true.

In general, arbitrary nonconvex optimization problems can be very challenging to solve. As a worst case example, consider a convex quadratic function $g(x) = x^2$ that is perturbed slightly: At some point where the function value is very large (say $x = 100$), we modify the function so that it suddenly drops to a large negative value (lower than the global minimum 0 of the convex quadratic). By doing this perturbation over a small finite interval , one can preserve differentiability while introducing a new global minimum far away from the original global minimum. In this way, one can create difficult optimization problems that cannot be solved using gradient descent methods, unless initialized very carefully.

The work we present here does not solve this problem: In fact, it will not find a global minimum of the form created above. The risk-aversion introduced destroys this minimum, since small perturbations cause the function to increase rapidly, ie, $g(\theta^* + \omega) \gg g(\theta^*)$, so that the objective (2) becomes large. Instead, it will find a "robust" minimum, in the sense that Gaussian perturbations around the minimum do not increase the value of the objective by much. This intuition is formalized by theorem 3.2, which bounds the suboptimality of the convexified solution relative to the optimal solution of the original problem in terms of the sensitivity of f to Gaussian perturbations around the optimum.

In figure 1, we illustrate the effect of the convexification for a 1-dimensional optimization problem. The blue curve represents the original function $g(\theta)$. It has 4-local minima in the interval $(-3, 3)$. Two of the shallow minima are eliminated by smoothing with Gaussian noise to get $\bar{g}(\theta)$. However, there is a deep but narrow local minimum that remains even after smoothing. Making the problem convex using risk-aversion and theorem 3.1 leads to the green curve that only preserves the robust minimum as the unique global optimum.

3.2 ANALYSIS OF SUB-OPTIMALITY

We have derived a convex surrogate for a very general class of optimization problems. However, it is possible that the solution to the convexified problem is drastically different from that of the original problem and the convex surrogate we propose is a poor 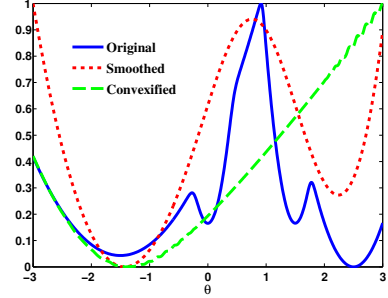approximation. Given the hardness of general non-convex optimization, we do not expect the two problems to have close solutions without additional assumptions. In this section, we analyze the gap between the original and convexified problem. In order to answer this, we first define the sensitivity function, which quantifies the gap between (1) and (2).

Figure 1: Illustration of Convexification for a 1-dimensional optimization problem

Definition 1 (Sensitivity Function)**.** The sensitivity function of f at noise level Σ, risk α is defined as:

$$
\begin{aligned}
\mathrm{S}_{\alpha,f}(\theta) &= \frac{1}{\alpha} \log \left(\mathrm{E}\left[\exp\left(\alpha\left(f(\theta + \omega) - \bar{f}(\theta) \right) \right) \right] \right) \\
&= \frac{1}{\alpha} \log \left(\mathrm{E}\left[\exp\left(\alpha \tilde{f}_\omega(\theta) \right) \right] \right).
\end{aligned}
$$

This is a measure of how sensitive f is to Gaussian perturbations at θ.

Theorem 3.2 (Suboptimality Analysis)**.** *Let $g(\theta) = f(\theta) + \frac{1}{2}\theta^T R\theta$ and define:*

$$
\theta_\alpha^* = \underset{\theta \in \mathcal{C}}{\mathsf{argmin}}\, f_\alpha(\theta) + \frac{1}{2}\theta^T R\theta, \quad \theta^* = \underset{\theta \in \mathcal{C}}{\mathsf{argmin}}\, \bar{g}(\theta).
$$

Then,

$$
\bar{g}(\theta_\alpha^*) - \bar{g}(\theta^*) \le \mathrm{S}_{\alpha,f}(\theta)
$$

Proof. We have that $\forall \theta$, $\bar{g}(\theta) = \bar{f}(\theta) + \frac{1}{2}\theta^T R\theta + \frac{1}{2}\mathrm{tr}(\Sigma R)$. From the convexity of the function $y \to \exp(\alpha y)$ and Jensen's inequality, we have

$$
\bar{g}(\theta_\alpha^*) = \bar{f}(\theta_\alpha^*) + \frac{1}{2}\theta_\alpha^{*T} R\theta_\alpha^* + \frac{1}{2}\mathrm{tr}(\Sigma R)
$$

$$
\le f_\alpha(\theta_\alpha^*) + \frac{1}{2}\theta_\alpha^{*T} R\theta_\alpha^* + \frac{1}{2}\mathrm{tr}(\Sigma R)
$$

Since $\theta_\alpha^* = \mathsf{argmin}_{\theta \in \mathcal{C}}\, f_\alpha(\theta) + \frac{1}{2}\theta^T R\theta$, for any $\theta \in \mathcal{C}$, we have

$$
\bar{g}(\theta_\alpha^*) \le f_\alpha(\theta) + \frac{1}{2}\theta^T R\theta + \frac{1}{2}\mathrm{tr}(\Sigma R)
$$

$$
= \mathrm{S}_{\alpha,f}(\theta) + \bar{g}(\theta).
$$

The result follows by plugging in $\theta = \theta^*$ and subtracting $\bar{g}(\theta^*)$ from both sides. □

Remark 1. Although we only prove suboptimality relative to the optimal solution of a smoothed version of g, we can extend the analysis to the optimal solution of g itself. Define $\theta^* = \mathsf{argmin}_{\theta \in \mathcal{C}}\, g(\theta)$. We can prove that

$$g\left(\theta_\alpha^*\right) - g\left(\theta^*\right) \leq \left(\bar{g}\left(\theta^*\right) - g\left(\theta^*\right)\right) + \left(g\left(\theta_\alpha^*\right) - \bar{g}\left(\theta_\alpha^*\right)\right) + \mathrm{S}_{\alpha,f}\left(\theta\right)$$

Assuming that g changes slowly around θ^* and θ_α^* (indicative of the fact that θ^* is a "robust" minimum and θ_α^* is the minimum of a robustified problem), we can bound the first term. We leave a precise analysis for future work.

3.2.1 BOUNDING THE SENSITIVITY FUNCTION

The sensitivity function is exactly the moment generating function of the 0-mean random variable $\tilde{f}_\omega(\theta)$. Several techniques have been developed for bounding moment generation functions in the field of concentration inequalities [Boucheron et al., 2013]. Using these techniques, we can bound the moment generating function (i.e. the sensitivity function) under the assumption that f is Lipschitz-continuous. Before stating the lemma, we state a classical result:

Theorem 3.3 (Log-Sobolev Inequality, [Boucheron et al., 2013], theorem 5.4). *Let* $\omega \sim \mathcal{N}\left(0, \sigma^2 I\right)$ *and* f *be any continuously differentiable function of* ω. *Then,*

$$\mathrm{E}\left[f\left(\omega\right)^2 \log\left(\frac{f\left(\omega\right)^2}{\mathrm{E}\left[f\left(\omega\right)^2\right]}\right)\right] \leq 2\sigma^2\, \mathrm{E}\left[\left\|\nabla f\left(\omega\right)\right\|^2\right].$$

Further, if f *is Lipschitz with Lipschitz constant* L, *then*

$$\mathrm{E}\left[\exp\left(\alpha\left(f\left(\omega\right) - \mathrm{E}\left[f\left(\omega\right)\right]\right)\right)\right] \leq \exp\left(\alpha^2 L^2 \sigma^2 / 2\right).$$

Finally, we can also prove that if f *is differentiable and* $\mathrm{E}\left[\exp \lambda \left\|\nabla f\left(\omega\right)\right\|^2\right] < \infty$ *for all* $\lambda < \lambda_0$, *then, given* η *such that* $\lambda\sigma^2 < \eta\lambda_0$ *and* $\lambda\eta < 2$, *we have*

$$\log\left(\mathrm{E}\left[\exp\left(\lambda\left(f\left(\omega\right) - \mathrm{E}\left[f\left(\omega\right)\right]\right)\right)\right]\right) \leq$$

$$\frac{\lambda\eta}{2 - \lambda\eta} \log\left(\mathrm{E}\left[\exp\left(\frac{\lambda\sigma^2}{\eta}\left\|\nabla f\left(\omega\right)\right\|^2\right)\right]\right)$$

Theorem 3.4 (Lipschitz Continuous Functions). *Assume the same setup as theorem 3.2. Suppose further that* f *is Lipschitz continuous with Lipschitz constant* L, *that is,*

$$\forall \theta, \theta', |f(\theta) - f(\theta')| \leq L\|\theta - \theta'\|.$$

Then, $\mathrm{S}_{\alpha,f}(\theta) \leq \frac{1}{2}\alpha L^2 \lambda_{\max}(\Sigma)$. *Further,*

$$\bar{g}\left(\theta\right) - \min_{\theta \in \mathcal{C}} \bar{g}\left(\theta\right) \leq \frac{\alpha L^2 \lambda_{\max}\left(\Sigma\right)}{2}.$$

Proof. We can write

$$\tilde{f}\left(\omega'\right) = f\left(\theta + \Sigma^{1/2}\omega'\right) - \bar{f}\left(\theta\right)$$

where $\omega' \sim \mathcal{N}\left(0, I_k\right)$.

$$\left|\tilde{f}\left(\omega_1'\right) - \tilde{f}\left(\omega_2'\right)\right| = \left|f\left(\theta + \Sigma^{1/2}\omega_1'\right) - f\left(\theta + \Sigma^{1/2}\omega_2'\right)\right|$$

$$\leq L\left\|\Sigma^{1/2}\left(\omega_1' - \omega_2'\right)\right\| \leq L\lambda_{\max}\left(\Sigma^{1/2}\right)\left\|\omega_1' - \omega_2'\right\|$$

Since $\lambda_{\max}\left(\Sigma^{1/2}\right) = \sqrt{\lambda_{\max}\left(\Sigma\right)}$, this shows that $\tilde{f}$ is Lipschitz with Lipschitz constant $\sqrt{\lambda_{\max}\left(\Sigma\right)}L$. The result now follows from theorem 3.3. From theorem 3.2, this implies that

$$\bar{g}\left(\theta\right) - \min_{\theta \in \mathcal{C}} \bar{g}\left(\theta\right) \leq \frac{\alpha L^2 \lambda_{\max}\left(\Sigma\right)}{2}.$$

$\square$

4 ALGORITHMS AND CONVERGENCE GUARANTEES

In general, the expectations involved in (2) cannot be computed analytically. Thus, we need to resort to sampling based approaches in order to solve these problems. This has been studied extensively in recent years in the context of machine learning, where stochastic gradient methods and variants have been shown to be efficient, particularly in the context training machine learning algorithms with huge amounts of data. We now describe stochastic gradient methods for solving (2) and adapt the convergence guarantees of stochastic gradient methods to our setting.

4.1 Stochastic Gradient Methods with Convergence Guarantees

In this section, we will derive gradients of the convex objective function (2). We will assume that the function f is differentiable at all $\theta \in \mathbf{R}^k$. In order to get unbiased gradient estimates, we exponentiate the objective (2) to get:

$$\mathbf{G}\left(\theta\right) = \mathop{\mathrm{E}}_{\omega \sim \mathcal{N}(0, \Sigma)}\left[\exp\left(\alpha f\left(\theta + \omega\right) + \frac{\alpha}{2}\theta^T R\theta\right)\right].$$

Since $f(\theta)$ is differentiable for all θ, so is $\exp\left(\alpha f\left(\theta + \omega\right) + \frac{1}{2}\theta^T\left(\alpha R\right)\theta\right)$. Further, suppose that

$$\mathrm{E}\left[\exp\left(2\alpha f\left(\theta + \omega\right)\right)\left\|\nabla f\left(\theta + \omega\right) + R\theta\right\|^2\right]$$

exists and is finite for each θ. Then, if we differentiate $\mathbf{G}(\theta)$ with respect to θ, we can interchange the expectation and differentiation to get

$$\mathrm{E}\left[\exp\left(\alpha f\left(\theta + \omega\right) + \frac{\alpha}{2}\theta^T R\theta\right)\left(\alpha\nabla f\left(\theta + \omega\right) + \alpha R\theta\right)\right]$$

Thus, we can sample $\omega \sim \mathcal{N}(0, \Sigma)$ and get an unbiased estimate of the gradient

$$\alpha \exp\left(\alpha f(\theta + \omega) + \frac{\alpha}{2}\theta^T R\theta\right)(\nabla f(\theta + \omega) + R\theta) \quad (5)$$

which we denote by $\hat{\nabla}\mathbf{G}(\theta, \omega)$. As in standard stochastic gradient methods, one saves on the complexity of a single iteration by using a single (or a small number of) samples to get a gradient estimate while still converging to the global optimum with high probability and in expectation, because over multiple iterations one moves along the negative gradient "on average".

Algorithm 1 Stochastic Gradient Method for (2)

$\theta \leftarrow 0$
for $i = 1, \ldots, T$ **do**
$\quad \omega \sim \mathcal{N}(0, \Sigma), \theta \leftarrow \mathrm{Proj}_{\mathcal{C}}\left(\theta - \eta_i \hat{\nabla}\mathbf{G}(\theta, \omega)\right)$
end for

From standard convergence theory for stochastic gradient methods [Bubeck, 2013], we have:

Corollary 1. *Suppose that* $\mathrm{E}\left[\left\|\hat{\nabla}\mathbf{G}(\theta, \omega)\right\|^2\right] \leq \zeta^2$, $\mathcal{C}$ *is contained in a ball of radius* $\mathrm{R}(\mathcal{C})$ *and* $\alpha R \succeq \Sigma^{-1}$. *Run algorithm 1 for* T *iterations with* $\eta_i = \frac{\mathrm{R}(\mathcal{C})}{\zeta}\sqrt{\frac{1}{2i}}$ *and define* $\hat{\theta} = \frac{1}{T}\sum_{i=1}^{T}\theta_i$, $\mathbf{G}^* = \mathrm{argmin}_{\theta \in \mathcal{C}} \mathbf{G}(\theta)$. *Then, we have*

$$\mathrm{E}\left[\mathbf{G}\left(\hat{\theta}\right)\right] - \mathbf{G}^* \leq \mathrm{R}(\mathcal{C})\zeta\sqrt{\frac{1}{2T}}$$

In the following theorem, we prove a convergence rate for algorithm 1.

Theorem 4.1. *Suppose that* $R = \kappa I, \Sigma = \sigma^2 I, \alpha\kappa \succeq \frac{1}{\sigma^2}$, $\mathcal{C}$ *is contained in a sphere of radius* $\mathrm{R}(\mathcal{C})$ *and that for all* $\theta \in \mathcal{C}$ $\bar{g}(\theta) \leq \overline{m}$. *Also, suppose that for some* $\beta < (\alpha)^{-1}$:

$$\frac{1}{2\alpha}\log\left(\underset{\omega \sim \mathcal{N}(0, \Sigma)}{\mathrm{E}}\left[\exp\left(2\frac{\alpha\sigma^2}{\beta}\|\nabla f(\theta + \omega)\|^2\right)\right]\right) \leq \gamma^2$$

Define $\delta = \sqrt{\frac{\beta\gamma^2}{\sigma^2(1-\alpha\beta)}} + \kappa\mathrm{R}(\mathcal{C})$. *Then, we can choose* $\zeta \leq \alpha^2\delta^2\exp\left(2\alpha(\overline{m} + \gamma^2) + \frac{\alpha\beta}{1-\alpha\beta} - \sigma^2\kappa\right)$.

Remark 2. The convergence guarantees are in terms of the exponentiated objective $\mathbf{G}(\theta)$. We can convert these into bounds on $\log(\mathbf{G}(\theta))$ as follows:

$$\mathrm{E}\left[\log\left(\mathbf{G}(\hat{\theta})\right)\right] \leq \log\left(\mathrm{E}\left[\mathbf{G}(\hat{\theta})\right]\right) \leq$$
$$\log\left(\mathbf{G}^* + \frac{\zeta\mathrm{R}(\mathcal{C})}{\sqrt{2T}}\right)$$

where the first inequality follows from concavity of the log function. Subtracting $\log(\mathbf{G}^*)$, we get

$$\mathrm{E}\left[\log\left(\mathbf{G}\left(\hat{\theta}\right)\right)\right] - \log(\mathbf{G}^*) \leq \log\left(1 + \frac{\zeta\mathrm{R}(\mathcal{C})}{\sqrt{2T}\mathbf{G}^*}\right).$$

5 CONTROL PROBLEMS

In this section, we extend the above approach to the control of discrete-time dynamical systems. Stochastic optimal control of nonlinear systems in general is a hard problem and the only known general approach is based on dynamic programming, which scales exponentially with the dimension of the state space. Algorithms that approximate the solution of the dynamic program directly (approximate dynamic programming) have been successful in various domains, but scaling these approaches to high dimensional continuous state control problems has been challenging. In this section, we pursue the alternate approach of policy search or policy gradient methods [Baxter and Bartlett, 2001]. These algorithms have the advantage that they are directly optimizing the performance of a control policy as opposed to a surrogate measure like the error in the solution to the Bellman equation. They have been used successfully for various applications and are closely related to path integral control [Kappen, 2005, Theodorou et al., 2010b]. However, in all of these approaches, there were no guarantees made regarding the optimality of the policy that the algorithm converges to (even in the limit of infinite sampling) or the rate of convergence.

In this work, we develop the *first* policy gradient algorithms that achieve the *globally* optimal solutions to a class of *risk-averse* policy optimization problems.

5.1 Problem Setup

We deal with arbitrary discrete-time dynamical systems of the form

$$\boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 & \ldots & \epsilon_{N-1} \end{pmatrix}^T \sim \mathbb{P}_{\epsilon}$$
$$x_1 = 0, \quad x_{t+1} = \mathcal{F}(x_t, y_t, \epsilon_t, t) \quad t = 1, \ldots, N-1 \tag{6}$$
$$y_t = u_t + \omega_t, \quad \omega_t \sim \mathcal{N}(0, \Sigma_t) \quad t = 1, \ldots, N-1 \tag{7}$$

where $x_t \in \mathbf{R}^{n_s}$ denotes the state, $y_t \in \mathbf{R}^{n_u}$ the effective control input, $\epsilon_t \in \mathbf{R}^p$ external disturbances, $u_t \in \mathbf{R}^{n_u}$ the actual control input, $\omega_t \in \mathbf{R}^{n_u}$ the control noise, $\mathcal{F} : \mathbf{R}^{n_s} \times \mathbf{R}^{n_u} \times \mathbf{R}^p \times \{1, \ldots, N-1\} \mapsto \mathbf{R}^{n_s}$ the discrete-time dynamics. In this section, we will use boldface to denote quantities stacked over time (like $\boldsymbol{\epsilon}$). Equation (6) can model any noisy discrete-time dynamical system, since $\mathcal{F}$ can be any function of the current state, control input and external disturbance

(noise). However, we require that all the control dimensions are affected by Gaussian noise as in (7). This can be thought of either as real actuator noise or artificial exploration noise. The choice of zero initial state $x_1 = 0$ is arbitrary - our results even extend to an arbitrary distribution over the initial state.

We will work with costs that are a combination of arbitrary state costs and quadratic control costs:

$$\sum_{t=1}^{N} \ell_t(x_t) + \sum_{t=0}^{N-1} \frac{u_t^T R_t u_t}{2} \qquad (8)$$

where $\ell_t(x_t)$ is the stage-wise state cost at time t. ℓ_t can be any bounded function of the state-vector x_t. Further, we will assume that the control-noise is non-degenerate, that is Σ_t is full rank for all $0 \leq t \leq N-1$. We denote $S_t = \Sigma_t^{-1}$. We seek to design feedback policies

$$u_t = K_t \phi(x_t, t), \phi : \mathbf{R}^{n_s} \times \{1, 2, \ldots, N-1\} \mapsto \mathbf{R}^r$$
$$K_t \in \mathbf{R}^{n_u \times r} \qquad (9)$$

to minimize the accumulated cost (8). We will assume that the features ϕ are fixed and we seek to optimize the policy parameters $\mathbf{K} = \{K_t : t = 1, 2, \ldots, N-1\}$. The stochastic optimal control problem we consider is defined as follows:

$$\underset{\mathbf{K}}{\text{Minimize}} \; \underset{\boldsymbol{\epsilon}, \omega_t}{\mathrm{E}} \left[\exp\left(\alpha \mathcal{L}(\mathbf{K})\right)\right]$$
$$\text{Subject to } x_1 = 0, x_{t+1} = \mathcal{F}(x_t, y_t, \epsilon_t, t)$$
$$y_t = u_t + \omega_t, u_t = \mathrm{K}_t \phi(x_t, t)$$
$$\boldsymbol{\epsilon} \sim \mathbb{P}_{\epsilon}, \omega_t \sim \mathcal{N}(0, \Sigma_t)$$
$$\mathcal{L}(\mathbf{K}) = \sum_{t=0}^{N} \ell_t(x_t) + \sum_{t=0}^{N-1} \frac{u_t^T R_t u_t}{2} \qquad (10)$$

This is exactly the same as the formulation in Risk Sensitive Markov Decision Processes [Marcus et al., 1997], the only change being that we have explicitly separated the noise appearing in the controls from the noise in the dynamical system overall. In this formulation, the objective depends not only on the average behavior of the control policy but also on variance and higher moments (the tails of the distribution of costs). This has been studied for linear systems under the name of LEQG control [Speyer et al., 1974]. α is called the risk factor: Large positive values of α result in strongly risk-averse policies while large negative values result in risk-seeking policies. In our formulation, we will need a certain minimum degree of risk-aversion for the resulting policy optimization problem to be convex.

5.2 Convex Controller Synthesis

Theorem 5.1. *If $\alpha R_t \succeq (\Sigma_t)^{-1} = S_t$ for $t = 1, \ldots, N-1$, then the optimization problem (10) is* convex.

Proof. We first show that for a fixed $\boldsymbol{\epsilon}$, the quantity $\mathrm{E}_{\omega_t \sim \mathcal{N}(0, \Sigma_t)}\left[\exp\left(\alpha \mathcal{L}(\mathbf{K})\right)\right]$ is a convex function of $\mathbf{K}$. Then, by the linearity of expectation, so is the original objective. We can write down the above expectation (omitting the normalizing constant of the Gaussian) as:

$$\int \exp\left(-\sum_{t=1}^{N-1} \frac{1}{2} \|y_t - K_t \phi(x_t, t)\|_{S_t}^2 + \alpha \mathcal{L}(\mathbf{K})\right) d\mathbf{y}$$

If we fix $\mathbf{y}, \boldsymbol{\epsilon}$, using (6), we can construct x_t for every $t = 1, \ldots, N$. Thus, $\mathbf{x}$ is a deterministic function of $\mathbf{y}, \boldsymbol{\epsilon}$ and does not depend on $\mathbf{K}$. The term inside the exponential can be written as

$$-\frac{1}{2} \left(\sum_{t=1}^{N-1} \|y_t\|_{S_t}^2\right) + \alpha \left(\sum_{t=1}^{N} \ell_t(x_t)\right)$$
$$+ \sum_{t=1}^{N-1} \frac{1}{2} \mathrm{tr}\left(\left(K_t^T(\alpha R_t - S_t) K_t\right) \phi(x_t, t) \phi(x_t, t)^T\right)$$
$$- \sum_{t=1}^{N-1} y_t^T S_t \phi(x_t, t) K_t$$

The terms on the first line don't depend on $\mathbf{K}$. The function $\left(K_t^T(\alpha R_t - S_t) K_t\right)$ is convex in $\mathbf{K}$ with respect to the semidefinite cone [Boyd and Vandenberghe, 2004] when $\alpha R_t - S_t \succeq 0$ and $\phi(x_t, t) \phi(x_t, t)^T$ is a positive semidefinite matrix. Hence the term on the second line is convex in $\mathbf{K}$. The term on the third line is linear in $\mathbf{K}$ and hence convex. Since exp is a convex and increasing function, the composed function (which is the integrand) is convex as well in $\mathbf{K}$. Thus, the integral is convex in $\mathbf{K}$. $\qquad \square$

We can add arbitrary convex constraints and penalties on $\mathbf{K}$ without affecting convexity.

Corollary 2. *The problem*

$$\min_{\mathbf{K}} \quad \underset{\boldsymbol{\epsilon} \sim \mathbb{P}_{\epsilon}, \omega_t \sim \mathcal{N}(0, \Sigma_t)}{\mathrm{E}} \left[\exp\left(\alpha \mathcal{L}(\mathbf{K})\right)\right]$$
$$\textit{Subject to} \qquad (6), (7), \mathbf{K} \in \mathcal{C} \qquad (11)$$

is a convex optimization problem for any arbitrary convex set $\mathcal{C} \subset \mathbf{R}^{n_u \times r \times (N-1)}$ if $\alpha R_t \succeq \Sigma_t \quad \forall t$.

6 NUMERICAL RESULTS

In this section, we present preliminary numerical results illustrating applications of the framework to various problems with comparisons to a simple baseline approach. These are not meant to be thorough numerical comparisons but simple illustrations of the power and applications of our framework.

6.1 BINARY CLASSIFICATION

We look at a problem of binary classification. Let y denote the actual label and $\hat{y}$ denote the predicted label. We use the loss function

$$\ell(y, \hat{y}) = \begin{cases} -\infty \text{ if } y\hat{y} > 0 \\ 0 \text{ otherwise} \end{cases}.$$

This is a non-convex loss function (the logarithm of the standard 0-1 loss). We convexify this in the prediction $\hat{y}$ using our approach:

$$\frac{1}{\alpha} \log \left(\mathrm{E}\left[\exp\left(\alpha\ell(y, \hat{y})\right)\right] \right) + \frac{(\hat{y})^2}{2\sigma^2}.$$

Plugging in $\hat{y} = \theta^T x$ where x is the feature vector, we get

$$\frac{1}{\alpha} \log \left(\mathop{\mathrm{E}}_{\omega \sim \mathcal{N}(0,\sigma^2)}\left[\exp\left(\alpha\ell(y, \theta^T x + \omega)\right)\right] \right) + \frac{(\theta^T x)^2}{2\alpha\sigma^2}.$$

Plugging in the expression for ℓ gives

$$\frac{1}{\alpha} \log \left(\frac{1}{2}\mathrm{erfc}\left(\frac{y\theta^T x}{\sqrt{2}\sigma} \right) \right) + \frac{(\theta^T x)^2}{2\alpha\sigma^2}.$$

where erfc is the Gaussian error function. Given a dataset $\{(x_i, y_u)\}$, we can form the empirical risk-minimization problem with this convexified objective:

$$\sum_{i=1}^{M} \frac{1}{\alpha} \log \left(\frac{1}{2}\mathrm{erfc}\left(\frac{y^i\theta^T x^i}{\sqrt{2}\sigma} \right) \right) + \frac{(\theta^T x^i)^2}{2\alpha\sigma^2}$$

We can drop the α since it only scales the objective (this is a consequence of the fact that $\exp(\ell)$ is 0-1 valued and does not change on raising it to a positive power). Thus, we finally end up with

$$\frac{1}{M} \sum_{i=1}^{M} \left(\log \left(\frac{1}{2}\mathrm{erfc}\left(\frac{y^i\theta^T x^i}{\sqrt{2}\sigma} \right) \right) + \frac{(\theta^T x^i)^2}{2\sigma^2} \right).$$

The first term is a data-fit term (a smoothed version of the 0-1 loss) and the second term is a regularizer. although we penalize the prediction $\theta^T x$ rather than θ itself. If x are normalized and span all directions, by summing over the entire dataset we get something close to the standard Tikhonov regularization.

We compare the performance of our convexification-based approach with a standard implementation of a Support Vector Machine (SVM) [Chang and Lin, 2011]. We use the breast cancer dataset from [Bache and Lichman, 2013]. We compare the two algorithms on various train-test splits of the dataset (without using cross-validation or parameter tuning). For each split, we create a noisy version of the dataset by

adding Gaussian noise to the labels and truncating to $+1/-1$: $\hat{y}^i = \mathrm{sign}\left(y^i + \omega\right), \omega \sim \mathcal{N}\left(0, \sigma^2\right)$. The accuracy of the learned classifiers on withheld test-data, averaged over 50 random train-test splits with label corruption as described above, are plotted as function of the noise level σ in figure 1. This is not a completely fair comparison since our approach explicitly optimizes for the worst case under Gaussian perturbations to the prediction (which can also be seen as a Gaussian perturbation to the label). However, as mentioned earlier, the purpose of these numerical experiments is to illustrate the applicability of our convexification approach to various problems so we do not do a careful comparison to robust variants of SVMs, which would be better suited to the setting described here.
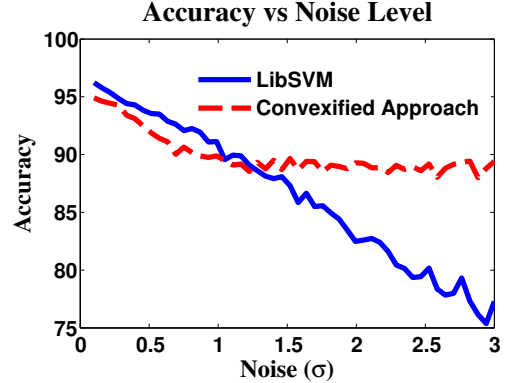

Figure 2: Binary Classification

6.2 CLASSIFICATION WITH NEURAL NETWORKS

We present an algorithm that does neural network training using the results of section 5. Each layer of the neural network is a time-step in a dynamical system, and the neural network weights correspond to the time-varying policy parameters. Let h denote a component-wise nonlinearity applied to its vector-input (a transfer function). The deterministic dynamics is

$$x_{t+1} = h\left(K_t x_t\right), x_0 = x$$

where x_t is the vector of activations at the t-th layer, K_t is the weight matrix and x is the input to the neural network. The output is x_N, where N is the number of layers in the network. The cost function is simply the loss function between the output of the neural network x_N and a desired output y: $\ell(y, x_N)$. To put this into our framework, we add noise to the input of the transfer function at each layer:

$$x_{t+1} = h\left(K_t x_t + \omega_t\right), x_1 = x, \omega_t \sim \mathcal{N}\left(0, \Sigma_t\right)$$

Additionally, we define the objective to be

$$
\mathrm{E}\left[\exp\left(\alpha\ell\left(y,x_N\right)+\sum_{t=1}^{N-1}\frac{\|K_t x_t\|_{\Sigma_t^{-1}}^2}{2}\right)\right]
$$

where the expectation is with respect to the Gaussian noise added at each layer in the network. Note that the above objective is a function of $\mathbf{K}, x, y$. The quadratic penalty on $K_t x_t$ can again be thought of as a particular type of regularization which encourages learning networks with small internal activations. We add this objective over the entire dataset $\{x^i, y^i\}$ to get our overall training objective.

We evaluate this approach on a small randomly selected subset of the MNIST dataset [LeCun et al., 1998]. We use the version available at http://nicolas.le-roux.name/ along with the MATLAB code provided for training neural networks. We use a 2-layer neural network with 20 units in the hidden layer and tanh-transfer functions in both layers. We use a randomly chosen collection of 900 data points for training and another 100 data points for validation. We compare training using our approach with simple backprop based training. Both of the approaches use a stochastic gradient - in our approach the stochasticity is both in selection of the data point i and the realization of the Gaussian noise ω while in standard backprop the stochasticity is only in the selection of i. We plot learning curves (in terms of generalization or test error) for both approaches, as a function of the number of neural network evaluations (forward+back prop) performed by the algorithm in figure 3a. The nonconvex approach based on standard backprop-gradient descent gets stuck in a local minimum and does not improve test accuracy much. On the other hand, the convexified approach is able to learn a classifier that generalizes better. We also compared backprop with training a neural network on a 1-dimensional regression problem where the red curve represents the original function with data-points indicated by squares, the blue curve the reconstruction learned by our convexified training approach and the black curve the reconstruction obtained by using backprop (figure 3b). Again, backprop gets stuck in a bad local minimum while our approach is able to find a fairly accurate reconstruction.

7   CONCLUSION AND FUTURE WORK

We have developed a general framework for convexifying a broad class of optimization problems, analysis that relates the solution of the convexified problem to the original one and given algorithms with convergence rate guarantees to solve the convexified problems. Ex-
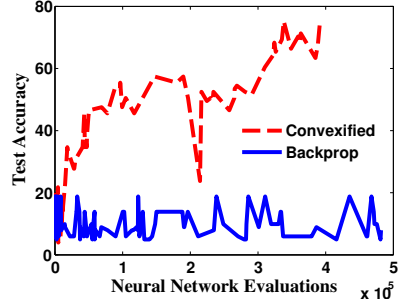
(a) Classification
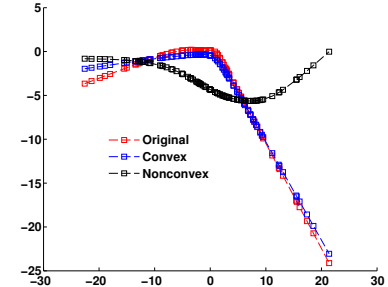
(b) Regression

Figure 3: Training Neural Networks

tending the framework to dynamical systems, we derive the first approach to policy optimization with optimality and convergence rate guarantees. We validated our approach numerically on problems of binary classification and training neural networks. In future work, we will refine the suboptimality analysis for our convexification approach. Algorithmically, stochastic gradient methods could be slow if the variance in the gradient estimates is high, which is the case when using the exponentiated objective (as in section 4). We will study the applicability of recent work on using better sampling algorithms with stochastic gradient [Atchade et al., 2014] to our convexified problems.

APPENDIX

Corollary 3. *[Boucheron et al., 2013], corollary 4.15 Let P, Q are arbitrary distributions on some space Ω and $f : \Omega \to \mathbf{R}$ is such that $\mathrm{E}_{\omega\sim P}\left[\exp\left(f\left(\omega\right)\right)\right] < \infty$. Then, the Kullback-Leibler divergence satisfies:*

$$
\mathrm{KL}\left(Q \parallel P\right) = \sup_{f}\left[\mathop{\mathrm{E}}_{\omega\sim Q}\left[f\left(\omega\right)\right] - \log\left(\mathop{\mathrm{E}}_{\omega\sim P}\left[\exp\left(f\left(\omega\right)\right)\right]\right)\right]
$$

7.1   PROOF OF THEOREM 4.1

Throughout this section, Let $\tilde{\alpha} = 2\alpha$ and P denote the Gaussian density $\mathcal{N}\left(0, \Sigma\right)$. Define a new distribution Q with density $Q\left(\omega\right) \propto P\left(\omega\right)\exp\left(\tilde{\alpha}\tilde{f}_{\omega}\left(\theta\right)\right)$. We denote $\hat{\nabla}\mathbf{G}\left(\theta, \omega\right)$ by $\hat{\nabla}\mathbf{G}$, $\nabla f\left(\theta + \omega\right)$ by $\hat{\nabla}f$ and

$\nabla f(\theta + \omega) + \kappa\theta$ by $\tilde{\nabla} f$ for brevity. Expectations are always with respect to $\omega \sim P$, unless denoted otherwise.

Proof. $(\alpha)^{-2} \, \mathrm{E}\left[\left\|\hat{\nabla}\mathbf{G}\right\|^2\right]$ evaluates to

$$
\exp\left(\alpha\kappa\theta^T\theta\right) \mathrm{E}\left[\exp\left(\tilde{\alpha} f(\omega + \theta)\right)\left\|\tilde{\nabla} f\right\|^2\right] =
$$

$$
\exp\left(\tilde{\alpha}\left(\frac{\kappa}{2}\theta^T\theta + \bar{f}(\theta)\right)\right) \mathrm{E}\left[\exp\left(\tilde{\alpha}\tilde{f}_\omega(\theta)\right)\left\|\tilde{\nabla} f\right\|^2\right] =
$$

$$
\exp\left(\tilde{\alpha}\left(\bar{g}(\theta) - \frac{1}{2}\sigma^2\kappa\right)\right) \mathrm{E}\left[\exp\left(\tilde{\alpha}\tilde{f}_\omega(\theta)\right)\left\|\tilde{\nabla} f\right\|^2\right] \tag{12}
$$

By the theorem hypotheses, the term outside the expectation is bounded above by $\alpha^2 \exp\left(2\alpha\overline{m} - \sigma^2\kappa\right)$. We are left with

$$
\mathrm{E}\left[\exp\left(\tilde{\alpha}\tilde{f}_\omega(\theta)\right)\left\|\tilde{\nabla} f\right\|^2\right].
$$

Dividing this by $\mathrm{E}\left[\exp\left(\tilde{\alpha}\tilde{f}_\omega(\theta)\right)\right]$, we get

$$
\frac{\mathrm{E}\left[\exp\left(\tilde{\alpha}\tilde{f}_\omega(\theta)\right)\left\|\tilde{\nabla} f\right\|^2\right]}{\mathrm{E}\left[\exp\left(\tilde{\alpha}\tilde{f}_\omega(\theta)\right)\right]} = \mathop{\mathrm{E}}_{\omega \sim Q}\left[\left\|\tilde{\nabla} f\right\|^2\right]. \tag{13}
$$

Expanding $\left\|\tilde{\nabla} f\right\|^2$, we get

$$
\left\|\hat{\nabla} f\right\|^2 + \kappa^2 \|\theta\|^2 + 2\hat{\nabla} f^T\theta
$$
$$
\leq \left\|\hat{\nabla} f\right\|^2 + \kappa^2 \mathrm{R}\left(\mathcal{C}\right)^2 + 2\kappa\left\|\hat{\nabla} f\right\|\mathrm{R}\left(\mathcal{C}\right)
$$

Finally from lemma 1, we have

$$
\mathop{\mathrm{E}}_{\omega \sim Q}\left[\left\|\hat{\nabla} f\right\|^2\right] \leq \frac{\beta\gamma^2}{\sigma^2(1 - \alpha\beta)}
$$

and by concavity of the square-root function,

$$
\mathop{\mathrm{E}}_{\omega \sim Q}\left[\left\|\hat{\nabla} f\right\|\right] \leq \sqrt{\frac{\beta\gamma^2}{\sigma^2(1 - \alpha\beta)}}.
$$

Plugging this bounds into the square expansion and letting $\delta = \sqrt{\frac{\beta\gamma^2}{\sigma^2(1 - \alpha\beta)}} + \kappa\mathrm{R}\left(\mathcal{C}\right)$, we get

$$
\mathop{\mathrm{E}}_{\omega \sim Q}\left[\left\|\nabla f(\theta + \omega) + \kappa\theta\right\|^2\right] \leq \delta^2 \tag{14}
$$

From (12),(13) and (14), $\mathrm{E}\left[\left\|\hat{\nabla}\mathbf{G}\right\|^2\right]$ is smaller than

$$
\alpha^2 \exp\left(2\alpha\overline{m} - \sigma^2\kappa\right)\delta^2 \mathrm{E}\left[\exp\left(\tilde{\alpha}\tilde{f}_\omega(\theta)\right)\right]
$$

Finally, by the last part of theorem 3.3,

$$
\mathrm{E}\left[\exp\left(\tilde{\alpha}\tilde{f}_\omega(\theta)\right)\right] \leq \exp\left(\frac{\alpha\beta}{1 - \alpha\beta} + 2\alpha\gamma^2\right).
$$

Combining the two above results gives the theorem. $\quad\square$

lemma 1. *Under the assumptions of theorem 4.1,*

$$
\mathop{\mathrm{E}}_{\omega \sim Q}\left[\sigma^2 \left\|\nabla f(\theta + \omega)\right\|^2\right] \leq \frac{\beta\gamma^2}{1 - \alpha\beta} \tag{15}
$$

Proof. The KL-divergence between Q and P is given by

$$
\int \frac{P(\omega)\exp\left(\tilde{\alpha}\tilde{f}_\omega(\theta)\right)}{\mathrm{E}\left[\exp\left(\tilde{\alpha}\tilde{f}_\omega(\theta)\right)\right]} \log\left(\frac{\exp\left(\tilde{\alpha}\tilde{f}_\omega(\theta)\right)}{\mathrm{E}\left[\exp\left(\tilde{\alpha}\tilde{f}_\omega(\theta)\right)\right]}\right)\mathrm{d}\,\omega.
$$

By theorem 3.3 applied to $\exp\left(\frac{\tilde{\alpha}}{2}f(\theta + \omega)\right)$, the above quantity is bounded above by $\frac{1}{2}(\tilde{\alpha}\sigma)^2 \mathrm{E}_{\omega \sim Q}\left[\left\|\hat{\nabla} f\right\|^2\right]$. Then, by corollary 3,

$$
\mathrm{KL}\left(Q \parallel P\right) \geq \mathop{\mathrm{E}}_{\omega \sim Q}\left[\frac{\tilde{\alpha}\sigma^2}{\beta}\left\|\hat{\nabla} f\right\|^2\right]
$$
$$
- \log\left(\mathrm{E}\left[\exp\left(\frac{\tilde{\alpha}\sigma^2}{\beta}\left\|\hat{\nabla} f\right\|^2\right)\right]\right).
$$

Denote the second term in the RHS by Γ. Plugging in the upper bound on $\mathrm{KL}\left(Q \parallel P\right)$, we get

$$
\Gamma \geq \sigma^2\left(\frac{\tilde{\alpha}}{\beta} - \frac{\tilde{\alpha}^2}{2}\right)\mathop{\mathrm{E}}_{\omega \sim Q}\left[\left\|\nabla f(\theta + \omega)\right\|^2\right].
$$

Since the LHS is upper bounded by $\tilde{\alpha}\gamma^2$ (hypothesis of theorem) which gives us the bound

$$
\mathop{\mathrm{E}}_{\omega \sim Q}\left[\sigma^2 \left\|\nabla f(\theta + \omega)\right\|^2\right] \leq \frac{2\beta\gamma^2}{2 - \tilde{\alpha}\beta} = \frac{\beta\gamma^2}{1 - \alpha\beta}.
$$

$\quad\square$

References

Y. F. Atchade, G. Fort, and E. Moulines. On stochastic proximal gradient algorithms. *ArXiv e-prints*, February 2014.

K. Bache and M. Lichman. UCI machine learning repository, 2013. URL http://archive.ics.uci.edu/ml.

J. Baxter and P. Bartlett. Infinite-horizon policy-gradient estimation. *Journal of Artificial Intelligence Research*, 15:319–350, 2001.

Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities: A nonasymptotic theory of independence.* Oxford University Press, 2013.

S. Boyd and L. Vandenberghe. *Convex optimization.* Cambridge University Press, Cambrdige, UK, 2004.

B. Van Den Broek, W. Wiegerinck, and B. Kappen. Risk sensitive path integral control. In *Uncertainty in AI, 2010. Proceedings of the 2010*, 2010.

Sebastian Bubeck. The complexities of optimization, December 2013. URL https://blogs.princeton.edu/imabandit/2013/04/25/orf523-noisy-oracles/.

Chih-Chung Chang and Chih-Jen Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2: 27:1–27:27, 2011. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

K. Dvijotham and E. Todorov. A unifying framework for linearly-solvable control. *Uncertainty in Artificial Intelligence*, 2011.

W. Fleming and S. Mitter. Optimal control and nonlinear filtering for nondegenerate diffusion processes. *Stochastics*, 8:226–261, 1982.

H.J. Kappen. Linear theory for control of nonlinear stochastic systems. *Physical Review Letters*, 95(20): 200201, 2005.

Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86 (11):2278–2324, 1998.

S.I. Marcus, E. Fernández-Gaucherand, D. Hernández-Hernandez, S. Coraluppi, and P. Fard. Risk sensitive Markov decision processes. *Systems and Control in the Twenty-First Century*, 29, 1997.

Hossein Mobahi. *Optimization by Gaussian Smoothing with Application to Geometric Alignment.* PhD thesis, University of Illinois at Urbana Champaign, Dec 2012.

A. Rangarajan. Generalised graduated non-convexity algorithm for maximum a posterjori image estimation. In *Proc. ICPR*, pages 127–133, 1990.

J Speyer, John Deyst, and D Jacobson. Optimization of stochastic linear systems with additive measurement and process noise using exponential per-

formance criteria. *Automatic Control, IEEE Transactions on*, 19(4):358–366, 1974.

E. Theodorou, J. Buchli, and S. Schaal. Reinforcement learning of motor skills in high dimensions: A path integral approach. In *Robotics and Automation (ICRA), 2010 IEEE International Conference on*, pages 2397–2403. IEEE, 2010a.

Evangelos Theodorou, Jonas Buchli, and Stefan Schaal. A generalized path integral control approach to reinforcement learning. *The Journal of Machine Learning Research*, 9999:3137–3181, 2010b.

E. Todorov. Efficient computation of optimal actions. *Proceedings of the National Academy of Sciences*, 106(28):11478, 2009.

M. Toussaint. Robot trajectory optimization using approximate inference. *International Conference on Machine Learning*, 26:1049–1056, 2009.