GeNGA: A Generalization of Natural Gradient Ascent with Positive and Negative Convergence Results

Philip S. Thomas PTHOMAS@CS.UMASS.EDU
School of Computer Science, University of Massachusetts, Amherst, MA 01003 USA

Abstract

Natural gradient ascent (NGA) is a popular optimization method that uses a positive definite metric tensor. In many applications the metric tensor is only guaranteed to be positive semidefinite (e.g., when using the Fisher information matrix as the metric tensor), in which case NGA is not applicable. In our first contribution, we derive *generalized natural gradient ascent* (GeNGA), a generalization of NGA which allows for positive semidefinite non-smooth metric tensors. In our second contribution we show that, in standard settings, GeNGA and NGA can both be divergent. We then establish sufficient conditions to ensure that both achieve various forms of convergence. In our third contribution we show how several reinforcement learning methods that use NGA without positive definite metric tensors can be adapted to properly use GeNGA.

1. Introduction

Natural gradient ascent (NGA) is a popular method for finding local maxima of a smooth function. According to Google Scholar, the paper introducing NGA (Amari, 1998) has over 1,600 citations from a wide range of fields, which hints at the popularity and impact of NGA. With its breadth of applications, it is not surprising that the assumptions made in the original derivation of NGA are not always satisfied. One example of this is that NGA assumes that the domain of the function being optimized is a Riemannian manifold. In some applications the domain is a semi-Riemannian manifold, but not necessarily a Riemannian manifold. Our first contribution is *generalized natural gradient ascent* (GeNGA), which relaxes the assumptions of NGA to allow for domains that are (possibly non-smooth) semi-Riemannian manifolds.

Despite the popularity of NGA, its convergence properties have not been well studied. In our second contribution, we provide positive and negative convergence results for NGA and GeNGA.

Finally, in our third contribution, we consider the application of NGA to reinforcement learning. This is an example of a field where the domain of the function to be optimized is a semi-Riemannian manifold, but not necessarily a Riemannian manifold. The derivations of the existing methods therefore, by using NGA, make an implicit assumption that is not satisfied. We remedy this by showing how reinforcement learning algorithms can use GeNGA in place of NGA.

The body of the paper is organized as follows. In Section 2 we show that NGA can diverge when ordinary gradient ascent converges. In Section 3 we derive GeNGA, and in Section 4 we provide convergence guarantees for NGA and GeNGA. In Section 5 we show how some reinforcement learning algorithms can be updated to use GeNGA. Finally, in Section 6 we present examples that illustrate the different assumptions and convergence guarantees before concluding in Section 7.

2. Divergence of Natural Gradient Ascent

Consider maximizing a function, $f : \mathbb{R}^n \to \mathbb{R}$.

Assumption 1. f is continuously differentiable. •

The gradient of f at x, $\nabla f(x)$, is the direction of change to x that causes f to increase most rapidly, assuming that x resides in Euclidean space. The natural gradient generalizes the gradient to allow x to lie on a Riemannian manifold with metric tensor $G(x)$ (Amari, 1998). On this manifold, at x, the length of a vector, $\Delta x \in \mathbb{R}^n$, is given by $\|\Delta x\|_{G(x)} := \sqrt{\Delta x^\intercal G(x) \Delta x}$.[1]

For G to describe a Riemannian manifold, it must vary smoothly from point to point and must be positive definite

Proceedings of the 31st International Conference on Machine Learning, Beijing, China, 2014. JMLR: W&CP volume 32. Copyright 2014 by the author(s).

[1] We use $A^\intercal$ to denote the transpose of a matrix, A, and A^+ to denote the Moore-Penrose pseudoinverse of A. When vector norms are applied to matrices, they denote the induced matrix norm. We assume that vectors are column vectors.

for all x, in which case $\|\cdot\|_{G(x)}$ are norms. Often, $G(x)$ is taken to be the Fisher information matrix for a parameterized distribution (Amari, 1998). NGA produces a sequence of points $(x_i)_{i=1}^{\infty}$ by ascending the natural gradient from an initial point, x_1, using a step size schedule $(\alpha_i)_{i=1}^{\infty}$ and the update

$$x_{i+1} = x_i + \alpha_i G(x_i)^{-1}\nabla f(x_i).$$

Despite claims that NGA converges to a local maximum (Peters & Schaal, 2008), without non-standard restrictions, it does not. Specifically, if Assumption 1 holds, and

Assumption 2 (Lipschitz Assumption). There exists a finite constant L such that $\forall x, z \in \mathbb{R}^n$,

$$\|\nabla f(x) - \nabla f(x-z)\|_2 \leq L\|z\|_2,$$

●

Assumption 3. All α_i are positive, $\sum_{i=1}^{\infty}\alpha_i = \infty$, and $\sum_{i=1}^{\infty}\alpha_i^2 < \infty$, ●

then *ordinary* gradient ascent causes either $f(x_i) \to \infty$ or $\lim_{i\to\infty}\|\nabla f(x_i)\|_2 = 0$ (Bertsekas & Tsitsiklis, 2000). However, the same is not true for NGA. In Theorem 1 we give an example where NGA oscillates and diverges when ordinary gradient ascent would converge to the global maximum.

Theorem 1. If Assumptions 1, 2, and 3 hold, then it can occur that the sequence, $(x_i)_{i=1}^{\infty}$, produced by NGA diverges when ordinary gradient ascent would converge to a finite value.

Proof. We provide a counterexample. Let $f(x) := -x^2$, where $x \in \mathbb{R}$. Notice that f is continuously differentiable and its derivative is Lipschitz. Consider the application of NGA to f with $\alpha_i = \frac{1}{2i}$ and $x_1 = 2$. Ordinary gradient ascent causes $x_i \to 0$ in this setting (Bertsekas & Tsitsiklis, 2000), and $x = 0$ is a global maximum of f. For NGA, let

$$G(x) := \begin{cases} -x^2 + 2 & \text{if } x \in (-1,1) \\ x^{-2} & \text{otherwise.} \end{cases}$$

This G meets all of the requirements to describe a Riemannian manifold.

When $x_i \notin (-1,1)$, the NGA update is

$$x_{i+1} = x_i + \alpha_i G(x_i)^{-1}\nabla f(x_i) = x_i - \frac{x_i^3}{i}.$$

We show that this sequence diverges without entering the $(-1,1)$ interval. We show this with an inductive proof that $|x_i| \geq 2i$.

We have two inductive hypotheses:

$$|x_i| \geq 2i, \quad (1)$$

$$\left|\frac{x_i^3}{i}\right| \geq 3|x_i|. \quad (2)$$

These are both satisfied when $i = 1$ since $x_1 = 2$.

For the inductive step for (1):

$$|x_{i+1}| = \left|x_i - \frac{x_i^3}{i}\right| \geq |2x_i|,$$

by (2). Then by (1): $|x_{i+1}| \geq |4i| > 2(i+1)$.

For the inductive step for (2):

$$\begin{aligned}\frac{|x_{i+1}|^3}{i+1} &= \frac{1}{i+1}\left|x_i - \frac{x_i^3}{i}\right|^2 |x_{i+1}| \\ &\geq \frac{1}{i+1}|2x_i|^2|x_{i+1}|,\end{aligned}$$

by (2). By (1) we have

$$\frac{|x_{i+1}|^3}{i+1} \geq \frac{1}{i+1}|4i|^2|x_{i+1}| \geq 3|x_{i+1}|.$$

□

In Section 6.5 we give an example to show that divergence can still occur even if the metric tensor is the Fisher information matrix. In the following section we introduce GeNGA, a generalization of NGA, before providing convergence proofs that apply to both methods.

3. Generalized Natural Gradient Ascent

Although the Fisher information matrix is often chosen as a metric tensor for NGA, it is only guaranteed to be positive *semi*definite. In these cases where $G(x)$ is only positive semidefinite, G describes a semi-Riemannian manifold and $\|\cdot\|_{G(x)}$ is a seminorm. Before deriving an expression for the directions of steepest ascent in this setting, we require:

Assumption 4. There exists at least one solution, Δx, to the equality

$$G(x)\Delta x = \nabla f(x),$$

for all $x \in \mathbb{R}^n$. ●

It can be shown that Assumption 4 implies that if there is a direction of change, Δx, to the current x that incurs no distance, then the directional derivative of f at x in the direction Δx is zero (i.e., $\nabla_{\Delta x} f(x) = 0$ for all x and Δx where $\|\Delta x\|_{G(x)} = 0$). We also use a similar but stronger assumption, which implies that $\nabla_{\Delta x} f(z) = 0$ for all x, z, and Δx where $\|\Delta x\|_{G(x)} = 0$:

Assumption 5. There exists at least one solution, Δx, to the equality $G(x)\Delta x = \nabla f(z)$, for all $x, z \in \mathbb{R}^n$. ●

In Theorem 2 we generalize the natural gradient to only require x to reside on a semi-Riemannian manifold.[2] Hereafter, to alleviate formatting problems, we write G for $G(x)$ and G_i for $G(x_i)$.[3] Also, let

$$h(x,v) := G^+\nabla f(x) + \left(I - G^+G\right) v, \tag{3}$$

where $v \in \mathbb{R}^n$.

Theorem 2. If each x lies on a semi-Riemannian manifold and Assumptions 1 and 4 hold, then for every v,

$$\frac{h(x,v)}{\|h(x,v)\|_G} \tag{4}$$

is a direction of steepest ascent of f at x. Also, every direction of steepest ascent is given by (4), for some v.

Proof. The directions of steepest ascent of f at x are the Δx that, for infinitesimal ϵ, maximize $f(x+\epsilon\Delta x)$, subject to $\|\Delta x\|_G = 1$ (Amari, 1998). By Assumption 1, the directions of steepest ascent are those that maximize $\epsilon\nabla f(x)^\intercal \Delta x$, subject to $\Delta x^\intercal G \Delta x - 1 = 0$. Using the method of Lagrange multipliers gives necessary conditions for Δx:

$$2\lambda G\Delta x = \nabla f(x), \tag{5}$$

for some positive scalar λ. The system of linear equations specified by (5) must have one solution (Assumption 4), and it may have many solutions since G is only positive semidefinite. Every solution is given by (4) for some v.

In the remainder of this proof we will show that $\nabla f(x)^\intercal h(x,v) / \|h(x,v)\|_G$ takes the same value for every v, and thus that all $h(x,v) / \|h(x,v)\|_G$ are directions of steepest ascent. This means that, in this instance, the method of Lagrange multipliers produces necessary *and sufficient* conditions.

By Assumption 4 we have that $GG^+\nabla f(x) = \nabla f(x)$. Using the definition of $h(x,v)$,

$$\begin{aligned} Gh(x,v) =& GG^+\nabla f(x) + \left[G - GG^+G\right] v \\ =& \nabla f(x). \end{aligned}$$

So,

$$\begin{aligned} \nabla f(x)^\intercal \frac{h(x,v)}{\|h(x,v)\|_G} =& \frac{\nabla f(x)^\intercal h(x,v)}{\sqrt{\nabla f(x)^\intercal h(x,v)}} \\ =& \sqrt{\nabla f(x)^\intercal h(x,v)} \\ =& \Big(\nabla f(x)^\intercal G^+ \nabla f(x) + \nabla f(x)^\intercal v \\ & - \nabla f(x)^\intercal G^+ G v\Big)^{\frac{1}{2}}. \end{aligned}$$

Since

$$\nabla f(x)^\intercal G^+ G v = \left(GG^+\nabla f(x)\right)^\intercal v = \nabla f(x)^\intercal v,$$

we have that

$$\nabla f(x)^\intercal \frac{h(x,v)}{\|h(x,v)\|_G} = \sqrt{\nabla f(x)^\intercal G^+ \nabla f(x)}. \tag{6}$$

Since the right side of (6) does not depend on v, all v cause $\nabla f(x)^\intercal h(x,v) / \|h(x,v)\|_G$ to take the same value. □

Given some $x_1 \in \mathbb{R}^n$, *generalized natural gradient ascent* (GeNGA) produces a sequence, $(x_i)_{i=1}^\infty$, by

$$x_{i+1} = x_i + \alpha_i \widetilde{\nabla} f(x_i),$$

where $(\alpha_i)_{i=1}^\infty$ is a sequence of non-negative step sizes, and where the *generalized natural gradient*, $\widetilde{\nabla} f(x_i)$, points in a direction of steepest ascent (but is not normalized):

$$\widetilde{\nabla} f(x) := h(x,v), \tag{7}$$

for some v. When G is positive definite for all x, this degenerates to NGA. Also, notice that, from the semi-Riemannian point of view (measuring distances using $\|\cdot\|_G$ rather than $\|\cdot\|_2$), the length of the generalized natural gradients are all equal:

$$\|\widetilde{\nabla} f(x)\|_G = \|G^+\nabla f(x)\|_G.$$

Lastly, selecting $v = 0$ gives the direction of steepest ascent with minimum Euclidean norm: $G^+\nabla f(x)$. We use an assumption to specify when we require GeNGA to use this direction of steepest ascent:

Assumption 6. $\widetilde{\nabla} f(x) = G^+\nabla f(x)$ always. •

4. Convergence

In this section we establish sufficient conditions to ensure that GeNGA achieves different types of convergence. We provide examples that illustrate the benefits and drawbacks of each convergence guarantee in Sections 6.1, 6.2, and 6.3.

One approach to showing that GeNGA converges is to match the requirements of an existing guarantee:

[2] We do not place smoothness restrictions on G for GeNGA or one of our convergence proofs, so this is actually more general than semi-Riemannian manifolds. However, to avoid convoluting the text, we still refer to x as residing on a semi-Riemannian manifold.

[3] This shorthand does not apply to any variables other than x. That is, G always denotes $G(x)$ and never $G(z)$. Sometimes we still write out $G(x)$ for emphasis.

Assumption 7. There exist positive scalars c_1 and c_2 such that

$$c_1 \left\|\nabla f(x_i)\right\|_2^2 \leq \nabla f(x_i)^\intercal \widetilde{\nabla} f(x_i), \tag{8}$$

and

$$\left\|\widetilde{\nabla} f(x_i)\right\|_2 \leq c_2 \left\|\nabla f(x_i)\right\|_2, \tag{9}$$

for all i. •

Theorem 3. If Assumptions 1, 2, 3, 4, and 7 hold, then GeNGA causes either $f(x_i) \to \infty$ or $\lim_{i\to\infty}\|\nabla f(x_i)\|_2 = 0$.

Proof. This follows immediately from the work of Bertsekas & Tsitsiklis (2000). □

A drawback of this guarantee is that different choices of v when G is singular (i.e., selecting different generalized natural gradients when there are many) can cause the left side of (9) to become arbitrarily large, which means that no c_2 can exist. This means that Theorem 3 is not always applicable to GeNGA. However, when using NGA or GeNGA with Assumption 6, Theorem 3 can be useful.

In the remainder of this section we present a guarantee that is more applicable to GeNGA. It uses less restrictive assumptions but provides a correspondingly weaker guarantee. In order to provide this guarantee, we introduce a modified Lipschitz assumption that uses the Riemannian seminorm in place of the Euclidean norm and a generalized natural gradient in place of the gradient:

Assumption 8 (Riemann-Lipschitz Assumption). There exists a finite constant, L_G, such that, $\forall x, z \in \mathbb{R}^n$,

$$\|G^+\nabla f(x) - G^+\nabla f(x-z)\|_G \leq L_G\|z\|_G. \quad \bullet$$

Intuitively, this says that, from a semi-Riemannian point of view, the gradient of f is Lipschitz. Notice that, for any x and A,

$$\|A^+x\|_A = \sqrt{x^\intercal A^+AA^+x} = \sqrt{x^\intercal A^+x} = \|x\|_{A^+}.$$

So, Assumption 8 implies that

$$\|\nabla f(x) - \nabla f(x-z)\|_{G(x)^+} \leq L_G\|z\|_{G(x)}. \tag{10}$$

We show in Theorem 4 that with different combinations of assumptions, GeNGA is guaranteed to converge to a desirable solution *from a semi-Riemannian point of view*, without the need for Assumption 7. That is, either $f(x_i) \to \infty$ or the magnitude of the generalized natural gradient (measured using the seminorm of the semi-Riemannian manifold) goes to zero:

Theorem 4. If either of the following sets of assumptions are satisfied:

1. Assumptions 1, 3, 5, and 8,
2. Assumptions 1, 3, 4, 6, and 8,

then the sequence, $(x_i)_{i=1}^\infty$ produced by GeNGA causes either $f(x_i) \to \infty$ or else $f(x_i)$ converges to a finite value and $\liminf_{i\to\infty}\|\widetilde{\nabla} f(x_i)\|_{G_i} = 0$.

Proof. We adapt a proof that ordinary gradient descent converges (Bertsekas & Tsitsiklis, 1997). For any $x, z \in \mathbb{R}^n$, let $g(\xi) := f(x - \xi z)$, where $\xi \in \mathbb{R}$. Then

$$\begin{aligned} f(x-z) - f(x) =& g(1) - g(0) \\ =& \int_0^1 \nabla g(\xi)\, \mathrm{d}\xi \\ =& -\int_0^1 z^\intercal \nabla f(x - \xi z)\, \mathrm{d}\xi. \end{aligned}$$

Adding $\int_0^1 z^\intercal(\nabla f(x) - \nabla f(x)) = 0$, we get

$$\begin{aligned} f(x-z) - f(x) =& -\int_0^1 z^\intercal \nabla f(x)\, \mathrm{d}\xi \\ & - \int_0^1 z^\intercal \left(\nabla f(x-\xi z) - \nabla f(x)\right) \mathrm{d}\xi \\ =& -z^\intercal\nabla f(x) - \int_0^1 z^\intercal \left(\nabla f(x-\xi z) - \nabla f(x)\right) \mathrm{d}\xi. \end{aligned}$$

Let $z := -\alpha_i h(x, v)$ and $x := x_i$. Then

$$\begin{aligned} f(x_i - z) - f(x_i) =& \alpha_i h(x,v)^\intercal \nabla f(x_i) \\ & - \int_0^1 \alpha_i h(x,v)^\intercal \left(\nabla f(x_i) - \nabla f(x_i - \xi z)\right) \mathrm{d}\xi \\ =& \alpha_i \left(G_i^+\nabla f(x_i) + \left[I - G_i^+G_i\right] v\right)^\intercal \nabla f(x_i) \\ & - \int_0^1 \alpha_i \left(G_i^+\nabla f(x_i) + \left[I - G_i^+G_i\right] v\right)^\intercal \\ & \left(\nabla f(x_i) - \nabla f(x_i - \xi z)\right) \mathrm{d}\xi \\ =& \alpha_i \nabla f(x_i)^\intercal G_i^+ \nabla f(x_i) \\ & + \alpha_i v^\intercal \left[I - G_iG_i^+\right] \nabla f(x_i) \qquad (11) \\ & - \int_0^1 \alpha_i \nabla f(x_i)^\intercal G_i^+ \left(\nabla f(x_i) - \nabla f(x_i - \xi z)\right) \mathrm{d}\xi \\ & - \int_0^1 \alpha_i v^\intercal \left[I - G_iG_i^+\right] \left(\nabla f(x_i) - \nabla f(x_i - \xi z)\right) \mathrm{d}\xi. \end{aligned} \tag{12}$$

If the first set of assumptions hold, then by Assumption 5, $GG^+\nabla f(x) = \nabla f(x)$ and $GG^+\nabla f(x - \xi z) = \nabla f(x - \xi z)$, so the terms on lines (11) and (12) are zero. If the second set of assumptions hold, then by Assumptions 4 and 6, $GG^+\nabla f(x) = \nabla f(x)$ and $v = 0$, so the terms on lines

(11) and (12) are zero. So, in both cases we get:

$$\begin{aligned} f(x_i - z) - f(x_i) \geq{}& \alpha_i \nabla f(x_i)^\intercal G_i^+ \nabla f(x_i) \\ &- \int_0^1 \alpha_i \nabla f(x_i)^\intercal G_i^+ \left(\nabla f(x_i) - \nabla f(x_i - \xi z)\right) \mathrm{d}\xi \\ \geq{}& \alpha_i \left\|\nabla f(x_i)\right\|^2_{G_i^+} \\ &- \int_0^1 \alpha_i \|\nabla f(x_i)\|_{G_i^+} \|\nabla f(x_i) - \nabla f(x_i - \xi z)\|_{G_i^+} \mathrm{d}\xi, \end{aligned}$$

by the Cauchy-Schwarz inequality for semi-inner-product spaces. By (10), which followed from Assumption 8,

$$\begin{aligned} f(x_i - z) - f(x_i) \geq{}& \alpha_i \left\|\nabla f(x_i)\right\|^2_{G_i^+} \\ &- \int_0^1 \alpha_i \|\nabla f(x_i)\|_{G_i^+} L_G \|\xi z\|_{G_i} \mathrm{d}\xi \\ ={}& \alpha_i \left\|\nabla f(x_i)\right\|^2_{G_i^+} \\ &- \int_0^1 \alpha_i^2 L_G \xi \|\nabla f(x_i)\|_{G_i^+} \|h(x,v)\|_{G_i} \mathrm{d}\xi. \end{aligned} \tag{13}$$

Notice that

$$\begin{aligned} \|h(x,v)\|^2_{G_i} ={}& h(x,v)^\intercal G_i h(x,v) \\ ={}& \left(G_i^+ \nabla f(x_i) + \left[I - G_i^+ G_i\right] v\right)^\intercal \\ & G_i \left(G_i^+ \nabla f(x_i) + \left[I - G_i^+ G_i\right] v\right) \\ ={}& \left(\nabla f(x_i)^\intercal G_i^+ + v^\intercal \left[I - G_i G_i^+\right]\right) \\ & G_i G_i^+ \nabla f(x_i) \\ ={}& \nabla f(x_i)^\intercal G_i^+ G_i G_i^+ \nabla f(x_i) \\ ={}& \nabla f(x_i)^\intercal G_i^+ \nabla f(x_i) \\ ={}& \|\nabla f(x_i)\|^2_{G_i^+}. \end{aligned}$$

So, continuing (13), we have

$$\begin{aligned} f(x_i - z) - f(x_i) \geq{}& \alpha_i \left\|\nabla f(x_i)\right\|^2_{G_i^+} \\ &- \int_0^1 \alpha_i^2 L_G \xi \|\nabla f(x_i)\|^2_{G_i^+} \mathrm{d}\xi \\ ={}& \alpha_i \left\|\nabla f(x_i)\right\|^2_{G_i^+} - \alpha_i^2 L_G \|\nabla f(x_i)\|^2_{G_i^+} \int_0^1 \xi \, \mathrm{d}\xi \\ ={}& \alpha_i \left\|\nabla f(x_i)\right\|^2_{G_i^+} - \frac{\alpha_i^2 L_G}{2} \|\nabla f(x_i)\|^2_{G_i^+}. \end{aligned}$$

So,

$$\begin{aligned} f(x_{i+1}) \geq{}& f(x_i) + \alpha_i \left(1 - \frac{\alpha_i L_G}{2}\right) \|\nabla f(x_i)\|^2_{G_i^+} \\ ={}& f(x_i) + \alpha_i \left(1 - \frac{\alpha_i L_G}{2}\right) \|G_i^+ \nabla f(x_i)\|^2_{G_i}. \end{aligned}$$

Since $\alpha_i \to 0$, we have for some positive constant c and all i greater than some index $\bar{i}$,

$$f(x_{i+1}) \geq f(x_i) + \alpha_i c \|G_i^+ \nabla f(x_i)\|^2_{G_i}. \tag{14}$$

From this relation, we see that for $i \geq \bar{i}$, $f(x_i)$ is monotonically nondecreasing, so either $f(x_i) \to \infty$ or $f(x_i)$ converges to a finite value. If the former case holds we are done, so assume the latter case. By adding (14) over all $i \geq \bar{i}$, we obtain

$$c \sum_{i=\bar{i}}^{\infty} \alpha_i \|G_i^+ \nabla f(x_i)\|^2_{G_i} \leq \lim_{i\to\infty} f(x_i) - f(x_{\bar{i}}) < \infty.$$

We see that there cannot exist an $\epsilon > 0$ such that $\|G_i^+ \nabla f(x_i)\|^2_{G_i} > \epsilon$ for all i greater than some $\hat{i}$, since this would contradict the assumption $\sum_{i=0}^\infty \alpha_i = \infty$. Therefore, we must have $\liminf_{i\to\infty} \|G_i^+ \nabla f(x_i)\|_{G_i} = 0$. This implies our result since $\|\widetilde{\nabla} f(x_i)\|_{G(x_i)} = \|G_i^+ \nabla f(x_i)\|_{G_i}$. □

In some cases it can be challenging to show that the Riemann-Lipschitz assumption (Assumption 8) is satisfied. We therefore introduce a new assumption that can be used together with Assumption 2 to imply Assumption 8:

Assumption 9. There exists a positive scalar c_3 such that for all $x, z \in \mathbb{R}^n$, $\|z\|_2 \leq c_3 \|z\|_{G(x)}$. •

Notice that Assumption 9 can only be satisfied if G is always positive definite.

Lemma 1. Assumptions 2 and 9 imply Assumptions 6 and 8.

Proof. Assumption 9 implies that $G(x)$ is always positive definite, so Assumption 6 is satisfied. Next we show that Assumptions 2 and 9 imply Assumption 8.

$$\begin{aligned} & \|\nabla f(x) - \nabla f(x-z)\|^2_{G^+} \\ &= (\nabla f(x) - \nabla f(x-z))^\intercal G^+ (\nabla f(x) - \nabla f(x-z)) \\ &= \left\langle G^+(\nabla f(x) - \nabla f(x-z)), (\nabla f(x) - \nabla f(x-z)) \right\rangle_2 \\ &\leq \|G^+(\nabla f(x) - \nabla f(x-z))\|_2 \|\nabla f(x) - \nabla f(x-z)\|_2 \\ &\leq \|G^+(\nabla f(x) - \nabla f(x-z))\|_2 \|z\|_2 \qquad (15) \\ &\leq c_3^2 \|G^+(\nabla f(x) - \nabla f(x-z))\|_G \|z\|_G \qquad (16) \\ &= c_3^2 \|\nabla f(x) - \nabla f(x-z)\|_{G^+} \|z\|_G. \end{aligned}$$

where (15) comes from Assumption 2 and (16) comes from Assumption 9. Dividing both sides of the inequality by $\|\nabla f(x) - \nabla f(x-z)\|_{G^+}$, which is always positive, we have

$$\|\nabla f(x) - \nabla f(x-z)\|_{G^+} \leq c_3^2 \|z\|_G,$$

and hence Assumption 8 is satisfied with $L_G = c_3^2$. □

Notice that Assumptions 2 and 9 together are more restrictive than Assumption 8, so if they are not satisfied, it does not mean that Assumption 8 is also not satisfied.

Not only can we use Lemma 1 to replace Assumptions 6 and 8 with Assumptions 2 and 9 in the requirements for Theorem 4, but we can use Assumption 9 to provide a stronger guarantee:

Theorem 5. If Assumptions 1, 2, 3, 4, and 9 hold, then the sequence, $(x_i)_{i=1}^{\infty}$ produced by GeNGA causes either $f(x_i) \to \infty$ or else $f(x_i)$ converges to a finite value and $\liminf_{i\to\infty} \|\widetilde{\nabla} f(x_i)\|_2 = \liminf_{i\to\infty} \|\widetilde{\nabla} f(x_i)\|_{G_i} = 0$.

Proof. We have from Theorem 4 and Lemma 1 that either $f(x_i) \to \infty$ or $\liminf_{i\to\infty} \|\widetilde{\nabla} f(x_i)\|_{G_i} = 0$. If the former case holds we are done, so assume the latter case. By Assumption 9, $\liminf_{i\to\infty} \|\widetilde{\nabla} f(x_i)\|_{G_i} = 0$ implies $\liminf_{i\to\infty} \|\widetilde{\nabla} f(x_i)\|_2 = 0$. □

5. Generalized Natural Policy Gradient Methods

In the previous sections we introduced GeNGA and analyzed its convergence properties. In this section we consider a field, reinforcement learning, where NGA has been applied even though the domain of the function being optimized is not guaranteed to be a Riemannian manifold, but is guaranteed to be a semi-Riemannian manifold.

Reinforcement learning algorithms search for “good” policies for *Markov decision processes* (MDPs). Policies are distributions that are typically parameterized by a vector $\theta \in \mathbb{R}^n$. An objective function $J : \mathbb{R}^n \to \mathbb{R}$ is selected to capture the desired properties of good policies. Our results apply to the standard average-reward and discounted-reward objective functions (Sutton et al., 2000). Policy search algorithms search for θ that maximize J. (Natural) policy gradient algorithms estimate and ascend the (natural) gradient of J.

Natural policy gradient algorithms typically assume that $G(\theta)$ is positive definite and thus invertible. This is not the case for popular policy parameterizations like tabular softmax action selection when G is the (average) Fisher information matrix (see Section 6.4 for an example). In this section, we remove this assumption.

Some natural policy gradient algorithms estimate $G(\theta)$ and $\nabla J(\theta)$ and then select $\widetilde{\nabla} J(\theta) = G(\theta)^{-1}\nabla J(\theta)$ (Bhatnagar et al., 2009). This can be easily corrected to $\widetilde{\nabla} J(\theta) = G(\theta)^{+}\nabla J(\theta)$, which is a generalized natural gradient of J at θ. However, other natural policy gradient algorithms form estimates of the natural gradient directly, without estimating $G(\theta)$. We show that, without modification, they perform generalized natural gradient ascent.

We will consider $w \in \mathbb{R}^n$ that satisfy Assumption 10, which comes from the combination of a standard constraint (Sutton et al., 2000, Equation 3) with the standard definition of $G(\theta)$ (Kakade, 2002, Equation 2):

Assumption 10. $\nabla J(\theta) = G(\theta) w$. •

The *natural policy gradient theorem*, states that if w is chosen to satisfy Assumption 10, then $G(\theta)^{-1}\nabla J(\theta) = w$ (Kakade, 2002, Theorem 1). This is useful because accurate estimates of w that satisfy Assumption 10 can be formed from small amounts of data using temporal-difference learning algorithms (Peters & Schaal, 2008). However, this clearly requires that $G(\theta)$ is invertible.

Although NGA is not applicable when G is singular, GeNGA is. We extend the natural policy gradient theorem to allow for positive semidefinite $G(\theta)$. We find that every w that satisfies Assumption 10 is still a generalized natural gradient (unnormalized direction of steepest ascent) of J at θ.

Theorem 6 (Generalized Natural Policy Gradient Theorem). If w is selected such that Assumption 10 holds, then w is a generalized natural gradient of J at θ and every generalized natural gradient is given by a w that satisfies Assumption 10.

Proof. The w that satisfy Assumption 10 are all given by

$$w = G(\theta)^{+}\nabla J(\theta) + \left(I - G(\theta)^{+}G(\theta)\right) v,$$

for some v. Every solution to $G(\theta)w = \nabla J(\theta)$ is given by this equation for some v, and every v produces a solution. Notice from (3) that this is merely $h(\theta, v)$. By (7), these w are the generalized natural gradients. □

This means that natural policy gradient algorithms that use w that satisfy Assumption 10 as their steepest ascent directions are already implementing GeNGA, and will therefore work properly when $G(\theta)$ is positive semidefinite. This is important because most natural policy gradient algorithms work this way (Morimura et al., 2005; Peters & Schaal, 2008; Bhatnagar et al., 2009; Degris et al., 2012). Although the algorithms are correct, convergence proofs that assume that $G(\theta)$ is always positive definite (Bhatnagar et al., 2009) do not apply when $G(\theta)$ is only guaranteed to be positive semidefinite.

6. Examples

In this section we give examples to ground the preceding theory.

6.1. Convergence by Theorem 3

We present an example where Theorem 3 can be applied. Let $f(x) := -x^2$, $G(x) = 2 + \sin(x)$, $\alpha_i = \frac{1}{i}$, and x_1 be any finite value. It is straightforward to show that Assumptions 1, 2, 3, 4, and 7 hold. So by Theorem 3, GeNGA causes either $f(x_i) \to \infty$ or $\lim_{i\to\infty}\|\nabla f(x_i)\| = 0$. Since $f(x)$ is bounded above, only the latter can occur.

6.2. Convergence by Theorem 4

Consider the application of NGA to $f(x) = -(x^\intercal[1,-1]^\intercal)^2$, where $x \in \mathbb{R}$. Let $x_1 = [1,1]^\intercal$, $\alpha_i = \frac{1}{i}$, and

$$G(x) := \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}.$$

Notice that this G is positive semidefinite but not positive definite. At each step, there will be an infinite number of directions of steepest ascent. In some cases you cannot guarantee that $v = 0$ is selected, for example, when using the generalized natural policy gradient theorem (Theorem 6). For this example, we select $v = [t,t]^\intercal$.

Since (9) in Assumption 7 is not satisfied and Assumption 9 is not satisfied, Theorems 3 and 5 are not applicable. Similarly, Assumption 6 is not satisfied, so we cannot use the second set of requirements for Theorem 4. However, since Assumptions 1, 3, 5, and 8 are satisfied, by Theorem 4, the sequence, $(x_i)_{i=1}^\infty$ produced by NGA causes either $f(x_i) \to \infty$ or else $f(x_i)$ converges to a finite value and $\liminf_{i\to\infty}\|\widetilde{\nabla} f(x_i)\|_{G_i} = 0$. Since $f(x)$ is bounded above, only the latter can occur.

6.3. Convergence by Theorem 5

Consider the application of NGA to $f(x) = -\frac{1}{2}(x-10)^2$, where $x \in \mathbb{R}$. Let $x_1 = 1$, $\alpha_i = \frac{0.1}{i}$, and

$$G(x) := \begin{cases} \frac{1}{2-x} & \text{if } x \in [1,2) \\ 1 & \text{otherwise.} \end{cases}$$

This G is positive (definite) and Assumptions 1, 2, 3, 4, and 9 are satisfied. So, by Theorem 5, the sequence, $(x_i)_{i=1}^\infty$ produced by NGA causes either $f(x_i) \to \infty$ or else $f(x_i)$ converges to a finite value and $\liminf_{i\to\infty}\|\widetilde{\nabla} f(x_i)\|_2 = \liminf_{i\to\infty}\|\widetilde{\nabla} f(x_i)\|_{G_i} = 0$. Notice that (8) in Assumption 7 is not satisfied, so the requirements of Theorem 3 are not satisfied. This is an interesting example because it showcases the difference between the convergence guarantees of Theorems 3 and 5.

The GeNGA update when $x \in [1,2)$ is

$$x_{i+1} = x_i + \frac{x^2 - 12x + 20}{10i}.$$

It can be shown that this update causes $x_i \to 2$ without leaving the $[1,2)$ interval. This example is depicted in Figure 1. Notice that $\liminf_{i\to\infty}\|\widetilde{\nabla} f(x_i)\|_2 = \liminf_{i\to\infty}\|\widetilde{\nabla} f(x_i)\|_{G_i} = 0$, but neither $f(x_i) \to \infty$ nor $\liminf_{i\to\infty}\|\nabla f(x_i)\|_2 = 0$. That is, from a semi-Riemannian point of view, the algorithm did the correct thing—it moved an infinite distance towards the global maximum. However, from the Euclidean point of view, this only got it to 2.

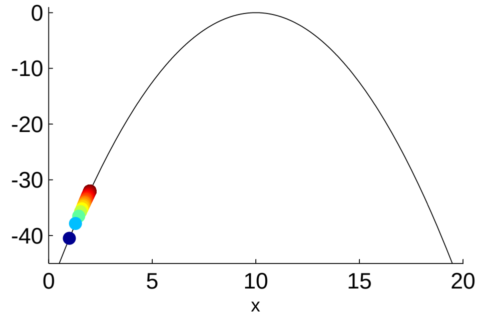

Figure 1. Sequence produced by NGA on the example from Section 6.3. The black curve is f and the points depict the sequence $(x_i)_{i=1}^\infty$, where x_1 is blue and x_i is dark red for large i. Notice that, from a Euclidean perspective, NGA converges prematurely.

6.4. Tabular Softmax Policies

Next, we present a simple reinforcement learning example where Assumption 5 is satisfied. This example also shows how the metric tensor can be positive semidefinite and not positive definite when using a common policy parameterization.

Consider any MDP with one state and two actions. We use a *softmax* policy parameterization. That is, the policy has parameter vector $\theta \in \mathbb{R}^2$ and the probability of action i is given by

$$\pi_\theta(i) := \frac{e^{\theta_i}}{e^{\theta_1} + e^{\theta_2}}.$$

Natural policy gradient methods typically use the average Fisher information matrix (Bagnell & Schneider, 2003) as their metric tensor. In this case:

$$\begin{aligned} G(\theta) :=& \sum_{i=1}^2 \pi_\theta(i) \frac{\partial \log \pi_\theta(i)}{\partial \theta} \frac{\partial \log \pi_\theta(i)}{\partial \theta}^\intercal \\ =& \pi_\theta(1)\pi_\theta(2) \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}. \end{aligned}$$

First, notice that $G(\theta)$ is positive semidefinite but not positive definite, and so GeNGA is required (in fact, a tabular softmax policy for any finite number of states and actions will result in $G(\theta)$ always being singular). Second, notice that the columns of $G(\theta)$ span all vectors in $\mathbb{R}^2$ that sum to zero.

This raises the question, why is it appropriate for $G(\theta)$ to not be positive definite? Allowing $G(\theta)$ to be positive semidefinite means that there are vectors, $\Delta\theta$, such that $\Delta\theta^\intercal G(\theta)\Delta\theta = \|\Delta\theta\|^2_{G(\theta)} = 0$. When computing the directions of steepest ascent, this means that there is a direction away from θ that does not incur any distance.

This is desirable when moving in the direction $\Delta\theta$ from θ does not change the distribution being optimized (in our case, the policy). Notice that we are parameterizing a distribution over two possible events. This should require only one parameter—the probability of one of the events, since the probability of the other is one minus this probability. However, our tabular softmax policy has two parameters. So, there are directions of change to the tabular softmax policy parameters that result in no change to the action probabilities (specifically, adding the same amount to both policy parameters). Moving along this direction should not incur distance when computing the directions of steepest ascent.

In this case, the gradient of the standard objective functions, J, at any θ', can both be written as (Sutton et al., 2000):

$$\nabla J(\theta') = \sum_{i=1}^{2} Q_{\theta'}(i)\frac{\partial \log \pi_{\theta'}(i)}{\partial \theta'},$$

where $Q_{\theta'}$ is a bounded real-valued function and θ' are any policy parameters. Since

$$\begin{aligned}\frac{\partial \log \pi_{\theta'}(1)}{\partial \theta'} =& \pi_{\theta'}(2)\begin{bmatrix}1\\-1\end{bmatrix}\\ \frac{\partial \log \pi_{\theta'}(2)}{\partial \theta'} =& \pi_{\theta'}(1)\begin{bmatrix}-1\\1\end{bmatrix},\end{aligned}$$

we have that $\nabla J(\theta')$ must be a vector in $\mathbb{R}^2$ that sums to zero, and hence it is in the column span of $G(\theta)$. So, Assumption 5 is satisfied.

6.5. Divergence of NGA for Policy Search

We showed in Theorem 1 that NGA can diverge. However, in that example we did not use the Fisher information matrix. This raises the question, can GeNGA, using the Fisher information matrix, diverge when optimizing a parameterized distribution, or does the Fisher information matrix introduce some properties that ensure convergence? We show that GeNGA can still diverge in this setting. In this example the Fisher information matrix is not always positive definite, so this example can not be used in place of the one used to prove Theorem 1.

Consider a bandit problem (one-state MDP with reward-discount parameter $\gamma = 0$) with two actions, a_1 and a_2. Let the reward for taking action a_1 be 1 and the reward for taking action a_2 be 0. In this setting, $J(\theta) = \Pr(a_1)$. We parameterize the policy with a single parameter, $\theta \in \mathbb{R}$, such that $\Pr(a_1|\theta) = f(\theta)$, where

$$f(\theta) := \begin{cases}\frac{1}{2\theta^2} + \frac{1}{2} & \text{if } \theta \notin [-2,2]\\ -\frac{\theta^2}{32} + \frac{3}{4} & \text{otherwise.}\end{cases}$$

$J(\theta) = f(\theta)$, so hereafter we discuss maximizing f. Also notice that, although f is defined in a piecewise manner, it is continuously differentiable and its derivative is Lipschitz. Let $\alpha_i = \frac{4}{i}$ and $\theta_1 = 5$. Ordinary gradient ascent on f causes $\theta_i \to 0$ in this setting (Bertsekas & Tsitsiklis, 2000), and $\theta = 0$ is a global maximum of f.

In this case, the Fisher information matrix can be written as

$$\begin{aligned}G(\theta) :=& \frac{\nabla f(\theta)^2}{f(\theta)\,(1-f(\theta))}\\ =& \begin{cases}\frac{4}{\theta^6-\theta^2} & \text{if } \theta \notin [-2,2]\\ \frac{-4\theta^2}{(\theta^2-24)(\theta^2+8)} & \text{otherwise.}\end{cases}\end{aligned}$$

So, when $\theta_i \notin [-2,2]$, the GeNGA update is

$$\begin{aligned}\theta_{i+1} =& \theta_i + \alpha_i G(\theta_i)^{-1}\nabla f(\theta_i)\\ =& \theta_i + \frac{1}{i\theta_i} - \frac{\theta_i^3}{i}.\end{aligned}$$

Since this sequence diverges without entering the $[-2,2]$ interval, we have that GeNGA causes $|\theta_i| \to \infty$ while ordinary gradient ascent causes $\theta_i \to 0$ (the global maximum).

7. Conclusion and Future Work

We presented GeNGA, a generalization of NGA to allow for positive semidefinite (possibly non-smooth) metric tensors. Next, we provided sufficient conditions to ensure that the sequences generated by GeNGA achieve different forms of convergence. We then showed how existing natural policy gradient algorithms could easily be updated to use GeNGA or already use GeNGA. Lastly, we provided examples to showcase the different types of convergence.

All of our convergence guarantees are for deterministic NGA and GeNGA. They do not apply when $\widetilde{\nabla} f(x)$ is not known, but noisy, biased estimates of it can be generated. It is straightforward to apply existing results in this setting if Assumption 7 holds (Bertsekas & Tsitsiklis, 2000). One avenue of future work would be to extend Theorems 4 and 5 to provide convergence guarantees in this setting, even when Assumption 7 does not hold.

References

Amari, S. Natural gradient works efficiently in learning. *Neural Computation*, 10:251–276, 1998.

Bagnell, J. A. and Schneider, J. Covariant policy search. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pp. 1019–1024, 2003.

Bertsekas, D. P. and Tsitsiklis, J. N. Gradient convergence in gradient methods. Technical Report LIDS-P ; 2404, Massachusetts Institute of Technology, Laboratory for Information and Decision Systems, November 1997.

Bertsekas, D. P. and Tsitsiklis, J. N. Gradient convergence in gradient methods with errors. *SIAM J. Optim.*, 10: 627–642, 2000.

Bhatnagar, S., Sutton, R. S., Ghavamzadeh, M., and Lee, M. Natural actor-critic algorithms. *Automatica*, 45(11): 2471–2482, 2009.

Degris, T., Pilarski, P. M., and Sutton, R. S. Model-free reinforcement learning with continuous action in practice. In *Proceedings of the 2012 American Control Conference*, 2012.

Kakade, S. A natural policy gradient. In *Advances in Neural Information Processing Systems*, volume 14, pp. 1531–1538, 2002.

Morimura, T., Uchibe, E., and Doya, K. Utilizing the natural gradient in temporal difference reinforcement learning with eligibility traces. In *International Symposium on Information Geometry and its Application*, 2005.

Peters, J. and Schaal, S. Natural actor-critic. *Neurocomputing*, 71:1180–1190, 2008.

Sutton, R. S., McAllester, D., Singh, S., and Mansour, Y. Policy gradient methods for reinforcement learning with function approximation. In *Advances in Neural Information Processing Systems 12*, pp. 1057–1063, 2000.