Journal of Machine Learning Research 17 (2016) 1-4 Submitted 10/16; Published 11/16

ERRATA
On the Estimation of the Gradient Lines of a Density and the Consistency of the Mean-Shift Algorithm

Ery Arias-Castro EARIASCA@MATH.UCSD.EDU
Department of Mathematics
University of California, San Diego
La Jolla, CA 92093, USA

David Mason DAVIDM@UDEL.EDU
Department of Applied Economics and Statistics
University of Delaware
Newark, DE 19717, USA

Bruno Pelletier BRUNO.PELLETIER@UNIV-RENNES2.FR
Département de Mathématiques
IRMAR – UMR CNRS 6625
Université Rennes II, France

Editor: Kevin Murphy

Our recent work (Arias-Castro et al., 2016) established the convergence of the mean shift algorithm under relatively general conditions. After the publication of this article, Prof. Jose E. Chacón — who has worked on the topic (Chacón and Monfort, 2013; Chacón and Duong, 2013) — alerted us of a mistake in the proof of the first part of our Theorem 1. The mistake is in the display following Eq. (32), where we applied the triangle inequality to the *squared* Euclidean distance, which of course is incorrect in general.

It turns out that the mistake has a simple and short fix, which we detail below. This relatively minor mistake would not warrant an errata, except that the same mistake has been made before by others also working on the convergence of the mean shift algorithm, including Comaniciu and Meer (2002), as revealed in (Li et al., 2007; Ghassabeh, 2015). (Prof. Chacón also provided these last two references.)

Below is a slightly modified statement of our Theorem 1 with the additional assumption that the end point $x^\star$ of the flow line is an isolated local maximum, meaning that for all $\epsilon > 0$ small enough $B(x^\star, \epsilon)$ contains no local maximum other than $x^\star$. Since f is also assumed of class C^3, this is equivalent to assuming that for all $\epsilon > 0$ small enough, $B(x^\star, \epsilon)$ contains only one critical point of f, namely $x^\star$, and $f(x^\star) > f(x)$ for all $x \in B(x^\star, \epsilon)$ such that $x \neq x^\star$.

The equations numbered (xx) refers to the original paper, while equations numbered (R.xx) are new to this note.

Theorem 1 *Let f be a function of class C^3. Let $(x(t) : t \geq 0)$ denote the flow line of f starting at x_0 and ending at an isolated local maximum $x^\star$ of f. Let (x_ℓ) be the sequence defined in (6) starting at x_0 . Then there exists $A = A(x_0, f) > 0$ such that, whenever*

$0 < a < A$,

$$\lim_{\ell \to +\infty} x_\ell = x^\star.$$

Denote by $x_a(t)$ *the following polygonal line*

$$x_a(t) = x_{\ell-1} + (t/a - \ell + 1)(x_\ell - x_{\ell-1}), \quad \forall t \in [(\ell-1)a, \ell a).$$

Assume $H_f(x^\star)$ *has all eigenvalues in* $(-\overline{\nu}, -\underline{\nu})$ *for some* $0 < \underline{\nu} < \overline{\nu}$. *Then, there exists a* $C = C(x_0, f, \underline{\nu}, \overline{\nu}) > 0$ *such that, for any* $0 < a < A$,

$$\sup_{t \geq 0} \|x_a(t) - x(t)\| \leq C a^\delta, \quad \delta := \frac{\underline{\nu}}{\underline{\nu} + \overline{\nu}}.$$

The rest of this note is dedicated to a corrected proof of the first part of this theorem. Claims 1 and 2 refer to the first two claims in the original published proof of Theorem 1. (Note that in the proof of Claim 1, we can assume without loss of generality that $f(x) > 0$ over $B(x_0, 3r_0)$. Also notice that x_0 is not a global minimum of f since x_0 is not a critical point of f and f is C^3.) In the published proof of Theorem 1, the second claim on page 15 should be removed.

Now replace the claim on page 16 by the following definition, observation and Claims A, B and C:

For any $\eta > 0$, denote by $\mathcal{C}(\eta)$ the connected component of $\mathcal{L}_f(f(x^\star) - \eta)$ that contains $x^\star$. Notice that since $x^\star$ is a local maximum, for all $\eta > 0$ small enough

$$\mathcal{C}(\eta) = \mathcal{C}(x^*, \eta), \tag{R.1}$$

where $\mathcal{C}(x^*, \eta)$ be the connected component of

$$\{y : f(x^*) - \eta \leq f(y) \leq f(x^*)\}$$

that contains x^*.

Claim A. *Let* y^* *be such that* $f(y^*) = f(x^*)$, *but* $y^* \neq x^*$. *For all* $\eta > 0$ *small enough* $y^* \notin \mathcal{C}(x^*, \eta)$. Choose such a y^*. Since x^* is an isolated local maximum, for all $\epsilon > 0$ small enough, $y^* \notin \overline{B}(x^*, \epsilon)$ and for some $\eta_\epsilon > 0$, $f(y) < f(x^*) - \eta_\epsilon$ for all $y \in \overline{B}(x^*, \epsilon) - B(x^*, \epsilon/2)$. Note that $\eta_\epsilon > 0$ can be chosen as small as desired by choosing $\epsilon > 0$ small enough. Suppose $y^* \in \mathcal{C}(x^*, \eta_\epsilon)$, then since $\mathcal{C}(x^*, \eta_\epsilon)$ is connected and $x^* \in \mathcal{C}(x^*, \eta_\epsilon)$ there is a continuous path lying inside $\mathcal{C}(x^*, \eta_\epsilon)$ joining x^* and y^*. Such a path would have to pass through a point $y \in \mathcal{C}(x^*, \eta_\epsilon) \cap \big(\overline{B}(x^*, \epsilon) - B(x^*, \epsilon/2)\big)$ for which $f(y) < f(x^*) - \eta_\epsilon$. This cannot happen, since $y \in \mathcal{C}(x^*, \eta_\epsilon)$ forces $f(x^*) - \eta_\epsilon \leq f(y)$. Hence for all for all $\eta > 0$ small enough $y^* \notin \mathcal{C}(x^*, \eta)$.

Claim B. *For all* $\eta > 0$ *small enough* $\mathcal{C}(x^*, \eta)$ *contains only one critical point of* f. Towards proving this we shall first show that for all $\epsilon > 0$ there exists an $\eta > 0$ such that

$$\mathcal{C}(x^*, \eta) \subset \overline{B}(x^*, \epsilon). \tag{R.2}$$

To see this, for any $\eta > 0$ small, denote the contour set

$$c(\eta) = \{y : f(y) = f(x^*) - \eta, y \in \mathcal{C}(x^*, \eta)\}.$$

Note that each contour set $c(\eta)$ is closed and hence is compact. (To see this, by Claim 1, we may assume that $\mathcal{L}(f(x_0))$ is bounded and thus compact. Therefore, since $c(\eta) \subset \mathcal{C}(x^*, \eta) \subset \mathcal{L}(f(x_0))$, $c(\eta)$ is compact.) Since $c\left(2^{-k}\right)$ is compact, for each $k \geq 1$ large enough there exists a $x_k \in c\left(2^{-k}\right)$ such that

$$r_k := \sup\left\{\|y - x^*\| : y \in c\left(2^{-k}\right)\right\} = \|x_k - x^*\|.$$

Observe that since $f(x_k) \to f(x^*)$ and x^* is isolated, necessarily by a compactness argument, $x_k \to x^*$. To see this, suppose that for some subsequence y_j of x_k we have $y_j \to y^*$. Necessarily $f(y^*) = f(x^*)$ and $y^* \in \mathcal{C}(x^*, \eta)$ for all $\eta > 0$. However by Claim A, necessarily $y^* = x^*$. Thus $x_k \to x^*$, which implies $r_k \to 0$. Noting that for all large k

$$\mathcal{C}\left(x^*, 2^{-k}\right) \subset \overline{B}(x^*, r_k),$$

we see that for all $\epsilon > 0$ there exists an $\eta > 0$ such that (R.2) holds. Therefore, since for all small enough $\epsilon > 0$, $\overline{B}(x^*, \epsilon)$ contains only one critical point of f, we get that for all $\eta > 0$ small enough $\mathcal{C}(x^*, \eta)$ contains only one critical point of f.

Claim C. (x_ℓ) *converges to* $x^\star$. By Claim B and (R.1), there exists $\eta_0 > 0$ small enough such that $\mathcal{C}(\eta_0)$ contains no critical point of f other than $x^\star$. Moreover since f is C^3, there exists $\epsilon > 0$ such that $\bar{B}(x^\star, \epsilon) \subset \mathcal{C}(\eta_0)$. Let ℓ_ϵ be such that $\|x(t_{\ell_\epsilon}) - x^\star\| \leq \epsilon/2$ and let a_ϵ be such that

$$\left[e^{\ell_\epsilon a_\epsilon \kappa_2 \sqrt{d}} - 1\right] \kappa_1 a_\epsilon = \epsilon/2.$$

Assume now that $a \leq A_1 \wedge a_\epsilon$, where A_1 is defined in (31). Then, by (33) and the triangle inequality,

$$\|x_{\ell_\epsilon} - x^\star\| \leq \|x_{\ell_\epsilon} - x(t_{\ell_\epsilon})\| + \|x(t_{\ell_\epsilon}) - x^\star\| \leq \epsilon.$$

Thus, x_{ℓ_ϵ} belongs to $\bar{B}(x^\star, \epsilon)$, and so to $\mathcal{C}(\eta_0)$. By Claim 2 the values of f are increasing along the polygonal curve x_a, so x_ℓ belongs to $\mathcal{C}(\eta_0)$ for all $\ell \geq \ell_\epsilon$.

Since the sequence $(f(x_\ell) : \ell \geq 0)$ is increasing and bounded, it is convergent and since

$$f(x_{\ell+1}) - f(x_\ell) \geq \frac{a}{2}\|\nabla f(x_\ell)\|^2,$$

we deduce that

$$\lim_{\ell \to \infty} \|\nabla f(x_\ell)\| = 0.$$

Recall that by Claim 1 we can assume that $\mathcal{L}(f(x_0))$ is bounded in which case $\mathcal{C}(\eta_0)$ is compact. Then we conclude that (x_ℓ) is convergent with $x_\ell \to x^\star$ by continuity of the gradient of f and the fact that $x^\star$ is the only critical point of f in $\mathcal{C}(\eta_0)$.

Acknowledgments

We are grateful to Jose E. Chacón for alerting us of the mistake and for referring us to the articles (Li et al., 2007; Ghassabeh, 2015).

References

Ery Arias-Castro, David Mason, and Bruno Pelletier. On the estimation of the gradient lines of a density and the consistency of the mean-shift algorithm. *Journal of Machine Learning Research*, 17(43):1–28, 2016.

José E Chacón and Tarn Duong. Data-driven density derivative estimation, with applications to nonparametric clustering and bump hunting. *Electronic Journal of Statistics*, 7: 499–532, 2013.

José E Chacón and Pablo Monfort. A comparison of bandwidth selectors for mean shift clustering. *arXiv preprint arXiv:1310.7855*, 2013.

D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):1–18, 2002.

Youness Aliyari Ghassabeh. A sufficient condition for the convergence of the mean shift algorithm with gaussian kernel. *Journal of Multivariate Analysis*, 135:1–10, 2015.

Xiangru Li, Zhanyi Hu, and Fuchao Wu. A note on the convergence of the mean shift. *Pattern Recognition*, 40(6):1756–1762, 2007.