On-Line Estimation with the Multivariate Gaussian Distribution

Sanjoy Dasgupta and Daniel Hsu

University of California, San Diego
{dasgupta,djhsu}@cs.ucsd.edu

Abstract. We consider on-line density estimation with the multivariate Gaussian distribution. In each of a sequence of trials, the learner must posit a mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$; the learner then receives an instance $\boldsymbol{x}$ and incurs loss equal to the negative log-likelihood of $\boldsymbol{x}$ under the Gaussian density parameterized by $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. We prove bounds on the regret for the follow-the-leader strategy, which amounts to choosing the sample mean and covariance of the previously seen data.

1 Introduction

We consider an on-line learning problem based on Gaussian density estimation in $\mathbb{R}^d$. The learning task proceeds in a sequence of trials. In trial t, the learner selects a mean $\boldsymbol{\mu}_t$ and covariance $\boldsymbol{\Sigma}_t$. Then, Nature reveals an instance $\boldsymbol{x}_t$ to the learner, and the learner incurs a loss $\ell_t(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t)$ equal to the negative log-likelihood of $\boldsymbol{x}_t$ under the Gaussian density parameterized by $(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t)$.

We will compare the total loss incurred from selecting the $(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t)$ in T trials to the total loss incurred using the best *fixed strategy* for the T trials. A fixed strategy is one that sets $(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t)$ to the same $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ for each t. The difference of these total losses is the *regret* of following a strategy and not instead selecting this best-in-hindsight $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ in every trial; it is the cost of not seeing all of the data ahead of time. In this paper, we will analyze the regret of the *follow-the-leader* strategy: the strategy which chooses $(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t)$ to be the sample mean and covariance of $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{t-1}\}$.

First, we find that a naïve formulation of the learning problem suffers from degenerate cases that lead to unbounded regret. We propose a straightforward alternative that avoids these problems by incorporating an additional, hallucinated, trial at time zero. In this setting, a trivial upper bound on the regret of follow-the-leader (FTL) is $\mathrm{O}(T^2)$ after T trials. We obtain the following bounds.

- For any $p > 1$, there are sequences $(\boldsymbol{x}_t)$ for which FTL has regret $\Omega(T^{1-1/p})$ after T trials. A similar result holds for any sublinear function of T.
- There is a linear bound on the regret of FTL that holds for all sequences.
- For any sequence, the average regret of FTL is ≤ 0 in the limit; formally,

$$\text{For any sequence } (\boldsymbol{x}_t),\ \limsup_{T\geq 1} \left\{ \frac{\text{Regret after } T \text{ trials}}{T} \right\} \leq 0.$$

N. Bshouty and C. Gentile (Eds.): COLT 2007, LNAI 4539, pp. 278–292, 2007.

On-line density estimation has been previously considered by Freund (1996), Azoury and Warmuth (2001), and Takimoto and Warmuth (2000a, 2000b). Collectively, they have considered the Bernoulli, Gamma, and fixed-covariance Gaussian distributions, as well as a general class of one-dimensional exponential families. However, on-line Gaussian density estimation with arbitrary covariance (that is, when the covariance is to be estimated) is all but unmentioned in the literature, even in the one-dimensional case. Indeed, these earlier bounds are logarithmic whereas most of ours are linear, a clear sign of a very different regime.

Learning a covariance matrix on-line is the main challenge not present in earlier analyses. Even in the univariate case, the total loss of the best fixed strategy in hindsight after T trials can lie anywhere in the range $[T - T\log T,\ T]$ (constants suppressed), while a learner that predicts a fixed variance $\sigma_t^2 \equiv c$ in every trial t will incur a total loss of at least $T \ln c$. This leaves the regret on the order of $T\log T$ in the worst case. Thus, even a linear regret bound is out of reach unless one makes an effort to estimate the variance.

Letting $\sigma^2(t)$ denote the sample variance of the first t observations, it turns out that our regret lower bounds are determined by sequences for which $\sigma^2(t) \to 0$ as t goes to infinity. On the other hand, if $\liminf \sigma^2(t) > 0$ – that is, if $\sigma^2(t)$ stays above a fixed constant for all $t > T_0$ – then it is easy to see from Lemmas 1 and 2 that the regret after T trials ($T > T_0$) is $\mathrm{O}(T_0 + \log(T/T_0))$. Thus, our results show that the performance of FTL depends on which of these two regimes the data falls under.

1.1 Related Work

On-line density estimation is a special case of sequential prediction with expert advice, a rich and widely applicable framework with roots in information theory, learning theory, and game theory (Cesa-Bianchi and Lugosi, 2006). In on-line density estimation, the set of experts is often an uncountably-infinite set, and the experts' predictions in a trial t only depend on the outcome $\boldsymbol{x}_t$ determined by Nature. Similar in spirit to density estimation is on-line subspace tracking (Crammer, 2006; Warmuth and Kuzmin, 2006). In the setup of Warmuth and Kuzmin, experts are low-dimensional linear subspaces, and the loss is the squared distance of $\boldsymbol{x}_t$ to the subspace (as in PCA).

We already mentioned work by Freund (1996), Azoury and Warmuth (2001), and Takimoto and Warmuth (2000a, 2000b). In each of the cases they considered, the regret bound is at most logarithmic in the number of trials. For the Bernoulli distribution, Freund showed that the Bayes algorithm with Jeffrey's prior asymptotically achieves the minimax regret. For the fixed-covariance Gaussian, Takimoto and Warmuth gave a recursively-defined strategy that achieves the minimax regret of $(r^2/2)(\ln T - \ln\ln T + \mathrm{O}(\ln\ln T/\ln T))$, where $\|\boldsymbol{x}_t\| \leq r$ for all $1 \leq t \leq T$.

Recent algorithms and frameworks for general on-line convex optimization (Zinkevich, 2003; Hazan et al, 2006; Shalev-Shwartz and Singer, 2006) are applicable to, among several other machine learning problems, many on-line density estimation tasks. However, they crucially rely on features of the loss function not enjoyed by the negative logarithm of the Gaussian density (e.g. finite minima, bounded derivatives). The follow-the-leader strategy and its variants are also applicable to many problems (Hannan, 1957; Kalai and Vempala, 2005; Zinkevich, 2003; Hazan et al, 2006). While FTL does not guarantee sublinear regret in many settings, several of the on-line density estimation algorithms derived by Azoury and Warmuth (2001) are special cases of FTL and do have logarithmic regret bounds.

2 On-Line Univariate Gaussian Density Estimation

To build intuition, we first demonstrate our results in the one-dimensional case before showing them in the general multivariate setting.

The learning protocol is as follows.

For trial $t = 1, 2, \ldots$

- The learner selects $\mu_t \in \mathbb{R}$ and $\sigma_t^2 \in \mathbb{R}_{>0} \triangleq \{x \in \mathbb{R} : x > 0\}$.
- Nature selects $x_t \in \mathbb{R}$ and reveals it to the learner.
- The learner incurs loss $\ell_t(\mu_t, \sigma_t^2)$ (ℓ_t implicitly depends on x_t).

The loss $\ell_t(\mu, \sigma^2)$ is the negative log-likelihood of x_t under the Gaussian density with mean μ and variance σ^2 (omitting the constant 2π),

$$\ell_t(\mu, \sigma^2) \;\triangleq\; -\ln \frac{1}{\sqrt{\sigma^2}} \exp\left\{ -\frac{(x_t - \mu)^2}{2\sigma^2} \right\} \;=\; \frac{(x_t - \mu)^2}{2\sigma^2} + \frac{1}{2} \ln \sigma^2 \qquad (1)$$

Suppose, over the course of the learning task, a strategy S prescribes the sequence of means and variances $((\mu_t, \sigma_t^2) : t = 1, 2, \ldots)$. We denote by $L_T(S)$ the total loss incurred by the learner following strategy S after T trials, and by $L_T(\mu, \sigma^2)$ the total loss incurred by the learner following the fixed strategy that selects $(\mu_t, \sigma_t^2) = (\mu, \sigma^2)$ for each trial t. So, we have

$$L_T(S) \;\triangleq\; \sum_{t=1}^{T} \ell_t(\mu_t, \sigma_t^2) \quad \text{and} \quad L_T(\mu, \sigma^2) \;\triangleq\; \sum_{t=1}^{T} \ell_t(\mu, \sigma^2). \qquad (2)$$

The learner seeks to adopt a strategy so that the regret after T trials

$$R_T(S) \;\triangleq\; L_T(S) \;-\; \inf_{\mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}_{>0}} L_T(\mu, \sigma^2) \qquad (3)$$

is as small as possible, even when Nature selects the x_t adversarially. Notice that, because $L_T(\mu, \sigma^2)$ is the likelihood of $\{x_1, x_2, \ldots, x_T\}$ under a single Gaussian model, the infimum in (3) is a maximum likelihood problem.

2.1 Degeneracies

Unfortunately, as the setting currently stands, the learner is doomed by two degeneracies that lead to unbounded regret. First, since we haven't restricted the magnitudes of the x_t, the regret can be unbounded even after just one trial. Takimoto and Warmuth (2000b) note that this is an issue even with fixed-variance Gaussian density estimation. Their remedy is to assume all $|x_t| \leq r$ for some $r \geq 0$, and we will do the same.

The second degeneracy is specific to allowing arbitrary variances and arises when the x_t are very close to each other. In fact, it stems from a standard difficulty with maximum likelihood estimation of Gaussians. To see the problem, suppose that the first few observations $x_1, x_2, \ldots, x_T$ are all the same. Then, while any reasonable learner must have set some nonzero variances σ_t^2 for $t = 1, 2, \ldots, T$ (for fear of facing an infinite penalty), the infimum of $L_T(\mu, \sigma^2)$ is $-\infty$ because the true variance of the data is 0. In fact, even if the x_t are not all the same, they can still be arbitrarily close together, leaving the infimum unbounded from below.

Our remedy is to hallucinate a zeroth trial that precludes the above degeneracy; it provides some small amount of variance, even if all the subsequent observations x_t are closely bunched together. Specifically, let $\tilde{\sigma}^2 > 0$ be some fixed constant. In the zeroth trial, we cause the learner to incur a loss of

$$\ell_0(\mu, \sigma^2) \triangleq \frac{1}{2} \sum_{x \in \{\pm\tilde{\sigma}\}} \left(\frac{(x-\mu)^2}{2\sigma^2} + \frac{1}{2} \ln \sigma^2 \right) = \frac{\mu^2 + \tilde{\sigma}^2}{2\sigma^2} + \frac{1}{2} \ln \sigma^2 .$$

Essentially, we hallucinate two instances, $\tilde{\sigma}$ and $-\tilde{\sigma}$, and incur half of the usual loss on each point.[1] This can be interpreted as assuming that there is some non-negligible variation in the sequence of instances, and for convenience, that it appears up front. We need to include the zeroth trial loss in the total loss after T trials. Thus, (2) should now read

$$L_T(S) \triangleq \sum_{t=0}^{T} \ell_t(\mu_t, \sigma_t^2) \quad \text{and} \quad L_T(\mu, \sigma^2) \triangleq \sum_{t=0}^{T} \ell_t(\mu, \sigma^2).$$

It can be shown that the infimum in (3) is always finite with the redefined $L_T(\mu, \sigma^2)$. With the extra zeroth trial, the infimum is no longer the Gaussian maximum likelihood problem; nevertheless, the form of the new optimization problem is similar. We have

$$\inf_{\mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}_{>0}} L_T(\mu, \sigma^2) = L_T(\overline{\mu}_T, \overline{\sigma}_T^2) = \frac{T+1}{2} + \frac{T+1}{2} \ln \overline{\sigma}_T^2 > -\infty \qquad (4)$$

for any $T \geq 0$, where

$$\overline{\mu}_T = \frac{1}{T+1} \sum_{t=1}^{T} x_t \quad \text{and} \quad \overline{\sigma}_T^2 = \frac{1}{T+1} \left(\tilde{\sigma}^2 + \sum_{t=1}^{T} x_t^2 \right) - \overline{\mu}^2 \geq \frac{\tilde{\sigma}^2}{T+1}$$

(the last inequality follows from Cauchy-Schwarz).

Before continuing, we pause to recap our notation and setting.

- $(\mu_t, \sigma_t^2) \in \mathbb{R} \times \mathbb{R}_{>0}$: parameters selected by the learner in trial $t \geq 0$.
- $x_t \in [-r, r]$: instances revealed to the learner in trial $t \geq 1$.

[1] Or, we take the expected loss of a zero-mean random variable with variance $\tilde{\sigma}^2$.

- $\ell_t(\mu, \sigma^2)$: loss incurred for selecting (μ, σ^2) in trial $t \geq 0$.
- $L_T(S)$: total loss of strategy S after T trials ($t = 1, 2, \ldots, T$), plus the loss incurred in the hallucinated zeroth trial ($t = 0$).
- $R_T(S) = L_T(S) - \inf_{(\mu,\sigma^2)} L_T(\mu, \sigma^2)$: regret after T trials of strategy S.

2.2 Follow-the-Leader

Motivated by the simplicity and success of the follow-the-leader based strategies for on-line density estimation with other distributions (Azoury and Warmuth, 2001), we instantiate such a strategy for on-line Gaussian density estimation. The name suggests using, in trial t, the setting of (μ, σ^2) that minimizes $L_{t-1}(\mu, \sigma^2)$. We will denote this setting as (μ_t, σ_t^2). It is precisely the values $(\overline{\mu}_{t-1}, \overline{\sigma}_{t-1}^2)$ given above; without the benefit of foresight, FTL is always one step behind the optimal strategy.

As noted in (Azoury and Warmuth, 2001), using FTL for on-line density estimation with exponential families leads to an intuitive recursive update. For the Gaussian distribution, it is

$$\mu_{t+1} = \mu_t + \frac{1}{t+1}(x_t - \mu_t) \quad \text{and} \quad \sigma_{t+1}^2 = \frac{t}{t+1}\sigma_t^2 + \frac{t}{(t+1)^2}(x_t - \mu_t)^2 \tag{5}$$

for $t \geq 1$. The loss function in the zeroth trial is fully known; so in the base cases, we have $(\mu_0, \sigma_0^2) = (0, \tilde{\sigma}^2)$ to optimize $\ell_0(\mu, \sigma^2)$, and $(\mu_1, \sigma_1^2) = (\mu_0, \sigma_0^2)$ as per FTL.

It will prove useful to derive an alternative expression for σ_t^2 by expanding the recursion in (5). We have $(t+1)\sigma_{t+1}^2 - t\sigma_t^2 = (t/(t+1)) \cdot (x_t - \mu_t)^2$ for $t \geq 1$; by telescoping,

$$\sigma_t^2 = \frac{1}{t}\left(\tilde{\sigma}^2 + \sum_{i=1}^{t-1} \Delta_i\right) \quad \text{where} \quad \Delta_t \triangleq \frac{t}{t+1}(x_t - \mu_t)^2. \tag{6}$$

2.3 Regret of Following the Leader

We obtain an expression for the regret $R_T \triangleq R_T(FTL)$ after T trials by analyzing the telescoping sum of $R_t - R_{t-1}$ from $t = 1$ to T. The difference $R_t - R_{t-1}$ is the *penalty* incurred by FTL for the additional trial t. The output our analysis will allow us to extract the core contribution of additional trials to the regret. Looking ahead, we'll show lower and upper bounds on the regret by focusing on this part of $R_t - R_{t-1}$.

Lemma 1. *The regret of FTL after T trials satisfies the bounds*

$$R_T \leq \sum_{t=1}^{T} \frac{1}{4(t+1)} \left[\frac{(x_t - \mu_t)^2}{\sigma_t^2}\right]^2 + \frac{1}{4}\ln(T+1) + \frac{1}{12} \quad \textit{and}$$

$$R_T \geq \sum_{t=1}^{T} \left(\frac{1}{4(t+1)} \left[\frac{(x_t - \mu_t)^2}{\sigma_t^2}\right]^2 - \frac{1}{6(t+1)^2} \left[\frac{(x_t - \mu_t)^2}{\sigma_t^2}\right]^3\right) + \frac{1}{4}\ln(T+1).$$

Proof. First, we make substitutions in $R_t - R_{t-1}$ using the FTL update rule (5) and the minimizer of $L_T(\mu, \sigma^2)$ (from (4)):

$$\begin{aligned}
R_t - R_{t-1} &= (L_t(FTL) - L_t(\mu_{t+1}, \sigma_{t+1}^2)) - (L_{t-1}(FTL) - L_{t-1}(\mu_t, \sigma_t^2)) \\
&= (L_t(FTL) - L_{t-1}(FTL)) + (L_{t-1}(\mu_t, \sigma_t^2) - L_t(\mu_{t+1}, \sigma_{t+1}^2)) \\
&= \left(\frac{(x_t - \mu_t)^2}{2\sigma_t^2} + \frac{1}{2}\ln\sigma_t^2\right) + \left(\frac{t}{2} + \frac{t}{2}\ln\sigma_t^2 - \frac{t+1}{2} - \frac{t+1}{2}\ln\sigma_{t+1}^2\right) \\
&= \frac{(x_t - \mu_t)^2}{2\sigma_t^2} - \frac{t+1}{2}\ln\frac{\sigma_{t+1}^2}{\sigma_t^2} - \frac{1}{2} \\
&= \frac{(x_t - \mu_t)^2}{2\sigma_t^2} - \frac{t+1}{2}\ln\left(\frac{t}{t+1} + \frac{t}{(t+1)^2}\cdot\frac{(x_t - \mu_t)^2}{\sigma_t^2}\right) - \frac{1}{2} \\
&= \frac{(x_t - \mu_t)^2}{2\sigma_t^2} - \frac{t+1}{2}\ln\left(1 + \frac{(x_t - \mu_t)^2}{(t+1)\sigma_t^2}\right) + \frac{t+1}{2}\ln\frac{t+1}{t} - \frac{1}{2}.
\end{aligned}$$

To deal with the first two summands, we employ Taylor expansions $z - z^2/2 + z^3/3 \geq \ln(1+z) \geq z - z^2/2$ for $z \geq 0$. To deal with the last two, we use Stirling's formula via Lemma 6 in the appendix (for a quick estimate, apply the same Taylor expansions). Finally, since the sum is telescoping and $R_0 = 0$, summing $R_t - R_{t-1}$ from $t = 1$ to T gives the bounds. $\square$

We let UB_t be the term inside the summation in the upper bound in Lemma 1, and LB_t be the corresponding term in the lower bound. Using the alternative expression for the variance (6), we get the following:

$$\sum_{t=1}^{T}\mathrm{LB}_t \leq R_T - \frac{1}{4}\ln(T+1) \leq \sum_{t=1}^{T}\mathrm{UB}_t + \frac{1}{12}$$

where

$$\mathrm{UB}_t \triangleq \frac{t+1}{4}\left[\frac{\Delta_t}{\tilde{\sigma}^2 + \sum_{i=1}^{t-1}\Delta_i}\right]^2 \quad \text{and} \quad \mathrm{LB}_t \triangleq \mathrm{UB}_t - \frac{t+1}{6}\left[\frac{\Delta_t}{\tilde{\sigma}^2 + \sum_{i=1}^{t-1}\Delta_i}\right]^3.$$

2.4 Lower Bounds

We exhibit a sequence (x_t) that forces the regret R_T incurred by FTL after T trials to be linear in T. The idea behind the sequence is to trick the learner into being "overly confident" about its choice of the mean μ_t and to then suddenly penalize it with an observation that is far from this mean. The initial ego-building sequence causes FTL to prescribe a σ_t^2 so small that when the penalty $(x_t - \mu_t)^2 \neq 0$ finally hits, the increase in regret $R_t - R_{t-1}$ is very large. In fact, this large increase in regret happens just once, in trial T.

To make this more precise, the form of LB_t suggests "choosing" $\Delta_t = 0$ for $1 \leq t \leq T-1$ and hitting the learner with $\Delta_T > 0$. Then, while $\mathrm{LB}_1 = \mathrm{LB}_2 = \cdots = \mathrm{LB}_{T-1} = 0$, the final contribution to the regret LB_T is linear in T. The necessary Δ_t are achieved with the sequence that has $x_1 = x_2 = \ldots = x_{T-1} = 0$ and $x_T = r$, so we get the following lower bound.

Theorem 1. *Suppose $r \leq \tilde{\sigma}$. For any $T \geq 1$, there exists a sequence (x_t) such that the regret of FTL after T trials is*

$$R_T \geq \frac{1}{12} \cdot \left(\frac{r}{\tilde{\sigma}}\right)^4 \cdot \frac{T^2}{T+1} + \frac{1}{4}\ln(T+1).$$

Proof. Using the sequence described above, we have $\Delta_T = T/(T+1)$ and all other $\Delta_t = 0$. By Lemma 1, substituting these values in LB_t gives the bound. □

While Theorem 1 says nothing about the regret after $T' > T$ trials, we can iterate the argument to give a sequence that forces FTL to incur nearly linear regret for infinitely many T. To motivate our argument, we first show one approach that doesn't work: namely, to keep penalizing the learner in successive trials after the one in trial T. That is, we set $\Delta_t = 0$ for $t < T$ and then $\Delta_T > 0$, $\Delta_{T+1} > 0$, $\Delta_{T+2} > 0$, and so on. The reason this is not too bad for the learner is that the denominator of LB_t increases significantly during $t = T+1, T+2, \ldots$; specifically, the denominator of LB_t increases quadratically, while the leading t only increases linearly. Eventually, the LB_t become more like $1/t$ instead of t.

Instead, we space out the non-zero penalties so that they strike only when FTL sets very small variances. Let $f : \mathbb{N} \to \mathbb{N}$ be an increasing function and f^{-1} be its inverse map. We will inflict the nth non-zero penalty in trial $f(n)$, so f can be thought of as the schedule of penalties. When f doles out the penalties sparingly enough, the regret after $f(n)$ trials is very close to being linear in $f(n)$.

Theorem 2. *Suppose $r \leq \tilde{\sigma}$. Let $f : \mathbb{N} \to \mathbb{N}$ be any increasing function and f^{-1} its inverse map. Then there exists a sequence (x_t) such that, for any T in the range of f, the regret of FTL after T trials is*

$$R_T \geq \frac{1}{6} \cdot \left(\frac{r}{\tilde{\sigma}}\right)^4 \cdot \frac{T+1}{(f^{-1}(T)+1)^2} + \frac{1}{4}\ln(T+1).$$

Proof. Following the discussion above, the sequence (x_t) is defined so that $\Delta_{f(n)} = r^2/2$ for all $n \geq 1$ and $\Delta_t = 0$ for all other t. Let $x_t = \mu_t - \mathrm{sign}(\mu_t) r\sqrt{(t+1)/(2t)}$ for t in the range of f, and $x_t = \mu_t$ elsewhere. In both cases, $|x_t| \leq r$. Then, in trial $f(n)$, we have

$$\begin{aligned}
\mathrm{LB}_{f(n)} &= \frac{f(n)+1}{4}\left[\frac{r^2/2}{\tilde{\sigma}^2 + (n-1)(r^2/2)}\right]^2 - \frac{f(n)+1}{6}\left[\frac{r^2/2}{\tilde{\sigma}^2 + (n-1)(r^2/2)}\right]^3 \\
&\geq \frac{f(n)+1}{6}\left[\frac{r^2/2}{\tilde{\sigma}^2 + (n-1)(r^2/2)}\right]^2 \\
&= \frac{f(n)+1}{6}\left(\frac{r^2/2}{\tilde{\sigma}^2/2}\right)^2 \left[\frac{1}{2 + (n-1)\frac{r^2/2}{\tilde{\sigma}^2/2}}\right]^2 \geq \frac{f(n)+1}{6}\left(\frac{r}{\tilde{\sigma}}\right)^4 \frac{1}{(n+1)^2}.
\end{aligned}$$

Then, Lemma 1 conservatively gives $R_{f(n)} \geq \mathrm{LB}_{f(n)} + (1/4)\ln(f(n)+1)$. □

If f is a polynomial of degree $p \geq 1$, we can actually sum (integrate) the LB_t from $t = 1$ to T (as opposed to just taking the final term LB_T) and yield a tighter bound $R_T \geq c \cdot (T+1)^{1-1/p} + (1/4)\ln(T+1)$ for some positive constant c. Notice that when $f(n) = \Theta(n)$ (the schedule used in our first attempt to give the bound), the bound has only the log term. Of course, there exists penalty schedules f for which $T/(f^{-1}(T))^2 = \omega(T^{1-1/p})$ for any $p \geq 1$. For example, if the penalty schedule is $f(n) = \Theta(\exp(n^2))$, then $T/(f^{-1}(T))^2$ is $\Omega(T/\log T)$.

2.5 Upper Bounds

We show two types of upper bounds on the regret of FTL. The first shows that the regret after T trials is at most linear in T. This bound is not immediately apparent from the Taylor approximation in Lemma 1: the σ_t^2 can be as small as $\tilde{\sigma}^2/(t+1)$, so each UB_t can be linear in t, which naïvely would give a *quadratic* upper bound on R_T. But this cannot be the case for all t: after all, σ_t^2 can only be very small in trial t if earlier trials have been relatively penalty-free. The key to the analysis is the potential function argument of Lemma 2, which shows that UB_t is at most a constant on average, and allows us to conclude the following.

Theorem 3. *For any $T \geq 1$ and any sequence (x_t), the regret of FTL after T trials is*

$$R_T \leq \frac{1}{4} \cdot \left(\left(\frac{2r}{\tilde{\sigma}}\right)^4 + \left(\frac{2r}{\tilde{\sigma}}\right)^2 \right) \cdot (T+1) + \frac{1}{4}\ln(T+1) + \frac{1}{12}.$$

Proof. We have $|\mu_t| \leq r$ since it is a convex combination of real numbers in $[-r, r]$. So $|x_t - \mu_t| \leq 2r$ by the triangle inequality; the theorem follows from combining Lemma 1 and Lemma 2 (below) with $c = (2r)^2/\tilde{\sigma}^2$, $a_1 = 0$, and $a_t = \Delta_{t-1}/\tilde{\sigma}^2$ for $2 \leq t \leq T+1$. □

Lemma 2. *For any $a_1, a_2, \ldots, a_T \in [0, c]$,*

$$\sum_{t=1}^{T} t \left[\frac{a_t}{1 + \sum_{i=1}^{t-1} a_i} \right]^2 \leq (c^2 + c) \cdot T \cdot \left(1 - \frac{1}{1 + \sum_{t=1}^{T} a_t} \right).$$

The bound in the lemma captures the fact that when $\sum_{i=1}^{t-1} a_i$ is small, a large penalty may be imminent, but when $\sum_{i=1}^{t-1} a_i$ is large, the tth penalty cannot be too large. The final parenthesized term $1 - 1/(1 + \sum_{i=1}^{T} a_i)$ is treated as 1 when we apply this lemma, but the more elaborate form is essential for the proof.

Proof. Trivial if $c = 0$. Otherwise, we proceed by induction on T. In the base case, we need to show $a_1^2 \leq (c^2 + c)(1 - 1/(1 + a_1))$; this follows because $a_1(1 + a_1) \leq c^2 + c$. For the inductive step, we assume the bound holds for $T - 1$ and show that it holds for T. Let $S_T = 1 + a_1 + \ldots a_{T-1}$. We need

$$(c^2 + c)(T-1)\left(1 - \frac{1}{S_T}\right) + T\left[\frac{a_T}{S_T}\right]^2 \leq (c^2 + c)T\left(1 - \frac{1}{S_T + a_T}\right).$$

After rearranging, this reads

$$1 + T\left(\frac{1}{S_T} - \frac{1}{S_T + a_T}\right) \geq \frac{1}{S_T} + \frac{T}{c^2 + c}\left[\frac{a_T}{S_T}\right]^2 .$$

Since $S_T \geq 1$ and $a_T \leq c$, we have $1 \geq 1/S_T$ and $1/S_T - 1/(S_T + a_T) \geq (a_T/S_T)^2/(c^2 + c)$, which suffices to give the required bound. □

The second upper bound we show concerns the *average (per-trial) regret*, R_T/T. This quantity reflects the improvement of a strategy over time; if R_T/T tends to a positive constant or worse, the strategy can be said to either stagnate or diminish over time.

Although Theorems 1 and 3 show that the worst-case regret of FTL after T trials is proportional to T, they don't imply that the average regret tends to a positive constant. Theorem 2 exhibits a sequence (x_t) for which the regret after T trials is nearly linear in T for infinitely many T, but the average regret still tends to 0. The following theorem complements this sublinear lower bound by showing that, indeed, the average regret of FTL is at most zero in the limit.

Theorem 4. *For any sequence (x_t), the average regret of FTL after T trials R_T/T satisfies*

$$\limsup_{T\geq 1} \frac{R_T}{T} \leq 0.$$

Proof. We'll show, for any $\varepsilon > 0$ sufficiently small, that $\limsup_{T\geq 1} R_T/T \leq \varepsilon$. The idea is to partition the trials into two sets: those in which $\Delta_t \leq b_\varepsilon$, for some constant b_ε (independent of T), and those in which $\Delta_t > b_\varepsilon$. The former trials produce small penalties: the constant b_ε is chosen so that the average of these penalties is at most ε. The latter set of trials have larger deviations-from-the-mean, but therefore cause the variance to rise substantially, which means they cannot contribute too heavily to regret. To analyze the trials in this second set, we consider the penalty schedule $f : \mathbb{N} \to \mathbb{N}$ such that the nth trial in this second set is $f(n)$. Because each $\Delta_{f(n)}$ is (relatively) large, we can show that, no matter the schedule f, the cumulative penalty from these trials is $o(T)$. This then implies that the average penalty is $o(1)$. The remaining terms in the regret are at most logarithmic in T, so they contribute $o(1)$ on average, as well.

We just need to detail our handling the penalties from the two sets of trials. Let $A \triangleq \{t \in \mathbb{N} : \Delta_t \leq b_\varepsilon\}$ and $B \triangleq \{t \in \mathbb{N} : \Delta_t > b_\varepsilon\}$, where $b_\varepsilon \triangleq \tilde{\sigma}^2(\sqrt{1+4\varepsilon}-1)/2$. Notice that $\tilde{\sigma}^2/b_\varepsilon \geq 1$ whenever $\varepsilon \leq 3/4$. Furthermore, let $A^t \triangleq A \cap \{1, 2, \ldots, t\}$ and $B^t \triangleq B \cap \{1, 2, \ldots, t\}$. By Lemma 2 and the choice of b_ε,

$$\frac{1}{T}\sum_{t\in A^T} \mathrm{UB}_t \leq \frac{1}{4}\left(\frac{b_\varepsilon^2}{\tilde{\sigma}^4} + \frac{b_\varepsilon}{\tilde{\sigma}^2}\right) + o(1) < \varepsilon + o(1).$$

If B is finite, then we're done. So assume B is infinite and index it with $\mathbb{N}$ by assigning the nth smallest element of B to $f(n)$. Define $f^{-1}(T) = \max\{n :$

$f(n) \le T\}$, so we have $f(f^{-1}(T)) \le T$ with equality when T is in the image of f. Then, using the fact $b_\varepsilon < \Delta_t \le (2r)^2$,

$$\begin{aligned}
\frac{1}{T}\sum_{t\in B^T} \mathrm{UB}_t &= \frac{1}{T}\sum_{t\in B^T} \frac{t+1}{4}\left[\frac{\Delta_t}{\tilde{\sigma}^2 + \sum_{i=1}^{t-1}\Delta_i}\right]^2 \\
&\le \frac{4r^4}{T}\sum_{t\in B^T}\frac{t+1}{(\tilde{\sigma}^2+\sum_{i\in B^{t-1}}\Delta_i)^2} \le \frac{4r^4}{T}\sum_{n=1}^{f^{-1}(T)}\frac{f(n)+1}{(\tilde{\sigma}^2+(n-1)b_\varepsilon)^2} \\
&\le \frac{4r^4}{Tb_\varepsilon^2}\sum_{n=1}^{f^{-1}(T)}\frac{f(n)+1}{n^2} \le \frac{4r^4}{Tb_\varepsilon^2}\left(\mathrm{o}(f(f^{-1}(T))) + \frac{\pi^2}{6}\right) = \mathrm{o}(1),
\end{aligned}$$

where the second-to-last step follows from Lemma 3. □

The following is a consequence of the fact that $\sum_{n\ge 1} 1/n^2$ is finite.

Lemma 3. *If $f : \mathbb{N} \to \mathbb{N}$ is strictly increasing, then $\sum_{k=1}^{n} f(k)/k^2 = \mathrm{o}(f(n))$.*

Proof. Fix any $\varepsilon > 0$, $n_0 \in \mathbb{N}$ such that $\sum_{k=n_0+1}^{\infty} 1/k^2 \le \varepsilon/2$, and $n_1 \in \mathbb{N}$ such that $f(n_0)/f(n_1) \le 3\varepsilon/\pi^2$. Then for any $n \ge n_1$,

$$\frac{1}{f(n)}\sum_{k=1}^{n}\frac{f(k)}{k^2} = \frac{1}{f(n)}\sum_{k=1}^{n_0}\frac{f(k)}{k^2} + \frac{1}{f(n)}\sum_{k=n_0+1}^{n}\frac{f(k)}{k^2} \le \frac{f(n_0)}{f(n)}\sum_{k=1}^{n_0}\frac{1}{k^2} + \sum_{k=n_0+1}^{n}\frac{1}{k^2}$$

which, by the choices of n_0 and n_1, is at most ε. □

3 On-Line Multivariate Gaussian Density Estimation

In the d-dimensional setting, the learning protocol is generalized to the following.

For trial $t = 1, 2, \ldots$

- The learner selects $\boldsymbol{\mu}_t \in \mathbb{R}^d$ and $\boldsymbol{\Sigma}_t \in \mathbb{S}^d_{\succ \mathbf{0}} \triangleq \{\boldsymbol{X} \in \mathbb{R}^{d\times d} : \boldsymbol{X} = \boldsymbol{X}^\top, \boldsymbol{X} \succ \mathbf{0}\}$ (the cone of symmetric positive-definite $d \times d$ matrices).
- Nature selects $\boldsymbol{x}_t \in \mathbb{R}^d$ and reveals it to the learner.
- The learner incurs loss $\ell_t(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t)$.

The loss $\ell_t(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the negative log-likelihood of $\boldsymbol{x}_t$ under the multivariate Gaussian density with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ (omitting the $(2\pi)^d$),

$$\ell_t(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \;\triangleq\; \frac{1}{2}(\boldsymbol{x}_t - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_t - \boldsymbol{\mu}) + \frac{1}{2}\ln|\boldsymbol{\Sigma}|$$

where $|\boldsymbol{X}|$ denotes the determinant of a matrix $\boldsymbol{X}$.

3.1 Multivariate Degeneracies

Even in the case $d = 1$, we had to amend the setting to avoid trivial conclusions. Recall, the one-dimensional degeneracies occur when (1) the $|x_t|$ are unbounded, or (2) the x_t are all (nearly) the same. For arbitrary d, the first issue becomes unbounded $\|\boldsymbol{x}_t\|$; the remedy is to assume a bound $\|\boldsymbol{x}_t\| \leq r$ for all t. The second issue is similar to the one-dimensional case, except now the issue can occur along any dimension, such as when the $\boldsymbol{x}_t$ lie in (or are arbitrarily close to) a $k < d$ dimensional subspace. As before, we'll hallucinate a zeroth trial to preclude singularity in the data. For a known constant $\tilde{\sigma}^2 > 0$, the loss in this trial is

$$\ell_0(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \triangleq \mathbb{E}_{\boldsymbol{v}} \left(\frac{1}{2}(\boldsymbol{v} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{v} - \boldsymbol{\mu}) + \frac{1}{2} \ln |\boldsymbol{\Sigma}| \right)$$

where $\boldsymbol{v}$ is any zero-mean random vector with $\mathbb{E}\boldsymbol{v}\boldsymbol{v}^\top = \tilde{\sigma}^2 \boldsymbol{I}$ (for example, take $\boldsymbol{v}$ to be uniform over the the $2d$ points $\{\pm\tilde{\sigma}\sqrt{d}\boldsymbol{e}_i : i = 1, 2, \ldots, d\}$, where $\boldsymbol{e}_i$ is the ith elementary unit vector). The zeroth trial can be seen as assuming a minimal amount of full-dimensional variation in the data. Again, including the zeroth trial loss in the total loss is enough to ensure a non-trivial infimum of $L_T(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ over $\boldsymbol{\mu} \in \mathbb{R}^d$ and $\boldsymbol{\Sigma} \in \mathbb{S}^d_{\succ \mathbf{0}}$. We have

$$\inf_{\boldsymbol{\mu} \in \mathbb{R}^d, \boldsymbol{\Sigma} \in \mathbb{S}^d_{\succ \mathbf{0}}} L_T(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = L_T(\overline{\boldsymbol{\mu}}, \overline{\boldsymbol{\Sigma}}) = \frac{d(T+1)}{2} + \frac{T+1}{2} \ln |\overline{\boldsymbol{\Sigma}}| > -\infty \qquad (7)$$

for any $T \geq 0$, where

$$\overline{\boldsymbol{\mu}} = \frac{1}{T+1} \sum_{t=1}^{T} \boldsymbol{x}_t \quad \text{and} \quad \overline{\boldsymbol{\Sigma}} = \frac{1}{T+1} \left(\tilde{\sigma}^2 \boldsymbol{I} + \sum_{t=1}^{T} \boldsymbol{x}_t \boldsymbol{x}_t^\top \right) - \overline{\boldsymbol{\mu}}\,\overline{\boldsymbol{\mu}}^\top \succeq \frac{\tilde{\sigma}^2}{T+1} \boldsymbol{I} \succ \mathbf{0}.$$

3.2 Multivariate Follow-the-Leader and Regret Bounds

Follow-the-leader for multivariate Gaussian density estimation admits the following recursion for its setting of $(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t)$: for $t \geq 1$

$$\boldsymbol{\mu}_{t+1} = \boldsymbol{\mu}_t + \frac{1}{t+1}(\boldsymbol{x}_t - \boldsymbol{\mu}_t) \quad \text{and} \quad (t+1)\boldsymbol{\Sigma}_{t+1} = t\boldsymbol{\Sigma}_t + \boldsymbol{\Delta}_t \qquad (8)$$

where $\boldsymbol{\Delta}_t = (\boldsymbol{x}_t - \boldsymbol{\mu}_t)(\boldsymbol{x}_t - \boldsymbol{\mu}_t)^\top t/(t+1)$; the base cases are $(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) = (\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1) = (\mathbf{0}, \tilde{\sigma}^2 \boldsymbol{I})$.

Our bounds for FTL in the univariate case generalize to the following for the multivariate setting.

Theorem 5. *Suppose $r \leq \tilde{\sigma}$. For any $T \geq d$, there exists a sequence $(\boldsymbol{x}_t)$ such that the regret of FTL after T trials is*

$$R_T \geq \frac{d}{12} \cdot \left(\frac{r}{\tilde{\sigma}}\right)^4 \cdot \left(T - \frac{d}{2} + \frac{1}{2}\right) \left(\frac{T-d+1}{T-d+2}\right) \left(1 - \frac{d-1}{(T-d+1)(T-d+2)}\right)^2 + \frac{d}{4} \ln(T+1).$$

Theorem 6. *Suppose $r \leq \tilde{\sigma}$. For any strictly increasing $f : \mathbb{N} \to \mathbb{N}$ with $f(n) \geq dn$, there exists a sequence $(\boldsymbol{x}_t)$ such that, for any T in the range of f, the regret of FTL after T trials is*

$$R_T \geq \frac{d}{6} \cdot \left(\frac{r}{\tilde{\sigma}}\right)^4 \cdot \frac{T - (d/2) + (3/2)}{\left(f^{-1}(T) + 1\right)^2} + \frac{d}{4} \ln(T+1).$$

Theorem 7. *For any sequence $(\boldsymbol{x}_t)$ and any $T \geq 1$, the regret of FTL after T trials is*

$$R_T \leq \frac{d}{4} \cdot \left(\left(\frac{2r}{\tilde{\sigma}}\right)^4 + \left(\frac{2r}{\tilde{\sigma}}\right)^2\right) \cdot (T+1) + \frac{d}{4} \ln(T+1) + \frac{d}{12}.$$

Theorem 8. *For any sequence $(\boldsymbol{x}_t)$, the average regret of FTL after T trials R_T/T satisfies $\limsup_{T \geq 1} R_T/T \leq 0$.*

We achieve the extra factor d in the lower bounds by using the sequences from the one-dimensional bound but repeating each non-zero penalty d times – one for each orthogonal direction. Some care must be taken to ensure that $\|\boldsymbol{x}_t\| \leq r$; also, the non-zero penalties are not all of the same value because they occur in different trials. For the upper bounds, the potential function has to account for variation in all directions; thus it is now based on $\mathrm{Tr}(\boldsymbol{\Sigma}_{T+1}^{-1})$ as opposed to the variance in any single direction.

3.3 Proof Sketches

We first need to characterize the penalty of FTL for each trial.

Lemma 4. *The regret of FTL after T trials satisfies the bounds*

$$R_T \leq \sum_{t=1}^{T} \frac{((\boldsymbol{x}_t - \boldsymbol{\mu}_t)^\top \boldsymbol{\Sigma}_t^{-1} (\boldsymbol{x}_t - \boldsymbol{\mu}_t))^2}{4(t+1)} + \frac{d}{4} \ln(T+1) + \frac{d}{12} \quad \textit{and}$$

$$R_T \geq \sum_{t=1}^{T} \left(\frac{((\boldsymbol{x}_t - \boldsymbol{\mu}_t)^\top \boldsymbol{\Sigma}_t^{-1} (\boldsymbol{x}_t - \boldsymbol{\mu}_t))^2}{4(t+1)} - \frac{((\boldsymbol{x}_t - \boldsymbol{\mu}_t)^\top \boldsymbol{\Sigma}_t^{-1} (\boldsymbol{x}_t - \boldsymbol{\mu}_t))^3}{6(t+1)^2} \right) + \frac{d}{4} \ln(T+1).$$

Proof. We proceed as in Lemma 1, using (8) and (7) to get

$$R_t - R_{t-1} = \frac{1}{2} (\boldsymbol{x}_t - \boldsymbol{\mu}_t)^\top \boldsymbol{\Sigma}_t^{-1} (\boldsymbol{x}_t - \boldsymbol{\mu}_t) - \frac{d}{2} + \frac{d(t+1)}{2} \ln \frac{t+1}{t} - \frac{t+1}{2} \ln \left| \boldsymbol{I} + \frac{1}{t+1} (\boldsymbol{x}_t - \boldsymbol{\mu}_t)(\boldsymbol{x}_t - \boldsymbol{\mu}_t)^\top \boldsymbol{\Sigma}_t^{-1} \right|.$$

The matrix inside the log-determinant has $d-1$ eigenvalues equal to 1 and one eigenvalue equal to $1 + (\boldsymbol{x}_t - \boldsymbol{\mu}_t)^\top \boldsymbol{\Sigma}_t^{-1} (\boldsymbol{x}_t - \boldsymbol{\mu}_t)/(t+1)$. Since the determinant of a matrix is the product of its eigenvalues, we can apply Taylor approximations $z - z^2/2 + z^3/3 \geq \ln(1+z) \geq z - z^2/2$ to the log-determinant, and Lemma 6 (in the appendix) to the other logarithm. □

Once again, we'll focus on the terms inside the summation. Let UB_t be the term under the summation in the upper bound, and LB_t be the that in the lower bound. Expanding the recursion for $\boldsymbol{\Sigma}_t$ in (8), we can express UB_t and LB_t as

$$\mathrm{UB}_t \triangleq (t+1)\mathrm{Tr}(\boldsymbol{\Delta}_t(\tilde{\sigma}^2\boldsymbol{I} + \textstyle\sum_{i=1}^{t-1}\boldsymbol{\Delta}_i)^{-1})^2/4 \quad \text{and}$$

$$\mathrm{LB}_t \triangleq \mathrm{UB}_t - (t+1)\mathrm{Tr}(\boldsymbol{\Delta}_t(\tilde{\sigma}^2\boldsymbol{I} + \textstyle\sum_{i=1}^{t-1}\boldsymbol{\Delta}_i)^{-1})^3/6$$

Lower Bounds. For Theorem 5, we want to cause non-zero penalties in orthogonal directions once the variance in these directions are small. The sequence begins with $\boldsymbol{x}_t = \boldsymbol{0}$ for $t \le T-d$, and for $i = 1, 2, \ldots, d$, has

$$\boldsymbol{x}_{T-d+i} = \boldsymbol{\mu}_{T-d+i} + r\sqrt{1 - \frac{\|\boldsymbol{\mu}_{T-d+i}\|^2}{r^2}}\boldsymbol{e}_i.$$

For Theorem 6, we combine the techniques from Theorem 5 and Theorem 2. Non-zero penalties occur in trials $f(n)-d+1, f(n)-d+2, \ldots, f(n)$ with $\|\boldsymbol{\delta}_t\|^2 = r^2/2$ in these trials and $\|\boldsymbol{\delta}_t\|^2 = 0$ in other trials.

Upper Bounds. The following generalization of Lemma 2 is the key argument for our upper bounds.

Lemma 5. *For any $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_T \in \mathbb{R}^d$ with $\|\boldsymbol{a}_t\|^2 \le c$,*

$$\sum_{t=1}^{T} t\mathrm{Tr}\left(\boldsymbol{A}_t\left(\boldsymbol{I} + \sum_{i=1}^{t-1}\boldsymbol{A}_i\right)^{-1}\right)^2 \le (c^2+c)\cdot T \cdot \left(d - \mathrm{Tr}\left(\left(\boldsymbol{I} + \sum_{i=1}^{T}\boldsymbol{A}_i\right)^{-1}\right)\right)$$

where $\boldsymbol{A}_i = \boldsymbol{a}_i\boldsymbol{a}_i^\top$ for all i.

Proof. Trivial if $c = 0$. Otherwise we proceed by induction on T. In the base case, we need $d(c^2+c) - (c^2+c)\mathrm{Tr}((\boldsymbol{I}+\boldsymbol{A}_1)^{-1}) - \|\boldsymbol{a}_1\|^4 \ge 0$. Using the Sherman-Morrison formula (for a matrix $\boldsymbol{M}$ and vector $\boldsymbol{v}$, $(\boldsymbol{M}+\boldsymbol{v}\boldsymbol{v}^\top)^{-1} = \boldsymbol{M}^{-1} - (\boldsymbol{M}^{-1}\boldsymbol{v}\boldsymbol{v}^\top\boldsymbol{M}^{-1})/(1+\boldsymbol{v}^\top\boldsymbol{M}^{-1}\boldsymbol{v})$), we have

$$(c^2+c)\mathrm{Tr}\left((\boldsymbol{I}+\boldsymbol{A}_1)^{-1}\right) = (c^2+c)\mathrm{Tr}\left(\boldsymbol{I} - \frac{\boldsymbol{A}_1}{1+\|\boldsymbol{a}_1\|^2}\right) = d(c^2+c) - \frac{(c^2+c)\|\boldsymbol{a}_1\|^2}{1+\|\boldsymbol{a}_1\|^2}$$

and also

$$\frac{(c^2+c)\|\boldsymbol{a}_1\|^2}{1+\|\boldsymbol{a}_1\|^2} - \|\boldsymbol{a}_1\|^4 \ge c\|\boldsymbol{a}_1\|^2 - \|\boldsymbol{a}_1\|^4 \ge 0.$$

Thus the base case follows. For the inductive step, we assume the bound holds for $T-1$ and show that it holds for T. Let $\boldsymbol{S} = \boldsymbol{I} + \boldsymbol{A}_1 + \ldots + \boldsymbol{A}_{T-1}$ and $\boldsymbol{A} = \boldsymbol{a}\boldsymbol{a}^\top = \boldsymbol{A}_T$. We need

$$(c^2+c)(T-1)\left(d - \mathrm{Tr}\left(\boldsymbol{S}^{-1}\right)\right) + T\mathrm{Tr}\left(\boldsymbol{A}\boldsymbol{S}^{-1}\right)^2 \le (c^2+c)T\left(d - \mathrm{Tr}\left((\boldsymbol{S}+\boldsymbol{A})^{-1}\right)\right),$$

which, after rearranging, reads

$$d + T\left(\mathrm{Tr}\left(\boldsymbol{S}^{-1}\right) - \mathrm{Tr}\left((\boldsymbol{S}+\boldsymbol{A})^{-1}\right)\right) \geq \mathrm{Tr}\left(\boldsymbol{S}^{-1}\right) + \frac{T\mathrm{Tr}\left(\boldsymbol{A}\boldsymbol{S}^{-1}\right)^2}{c^2+c}.$$

Since $\boldsymbol{S} \succeq \boldsymbol{I}$, we have $\mathrm{Tr}\left(\boldsymbol{S}^{-1}\right) \leq d$, which takes care of the first terms on each side. For the remaining terms, first note that $\mathrm{Tr}(\boldsymbol{A}\boldsymbol{S}^{-1}) \leq \|\boldsymbol{a}\|^2 \leq c$. Then, using Sherman-Morrison again gives

$$\mathrm{Tr}\left(\boldsymbol{S}^{-1}\right) - \mathrm{Tr}\left((\boldsymbol{S}+\boldsymbol{A})^{-1}\right) = \mathrm{Tr}\left(\frac{\boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S}^{-1}}{1+\mathrm{Tr}\left(\boldsymbol{A}\boldsymbol{S}^{-1}\right)}\right) = \frac{\|\boldsymbol{a}\|^2\boldsymbol{a}^\top\boldsymbol{S}^{-2}\boldsymbol{a}}{\|\boldsymbol{a}\|^2\left(1+\mathrm{Tr}\left(\boldsymbol{A}\boldsymbol{S}^{-1}\right)\right)}.$$

The denominator is at most $c(1+c)$, so it remains to show $\|\boldsymbol{a}\|^2\mathrm{Tr}(\boldsymbol{a}^\top\boldsymbol{S}^{-2}\boldsymbol{a}) \geq (\boldsymbol{a}^\top\boldsymbol{S}^{-1}\boldsymbol{a})^2$. Without loss of generality, $\|\boldsymbol{a}\| = 1$ and $\boldsymbol{S}$ is diagonal with eigenvalues $\lambda_1, \ldots, \lambda_d > 0$. Then $a_1^2/\lambda_1^2 + \ldots + a_d^2/\lambda_d^2 \geq (a_1^2/\lambda_1 + \ldots + a_d^2/\lambda_d)^2$ follows from Jensen's inequality. □

For Theorem 8, we proceed as in Theorem 4, but to handle the trials in B, we have to deal with each direction separately, so further partitions are needed.

4 Conclusion and Open Questions

On-line density estimation with a Gaussian distribution presents difficulties markedly different from those usually encountered in on-line learning. They appear even in the one-dimensional setting and scale up to the multivariate case as familiar issues in data analysis (e.g. unknown data scale, hidden low dimensional structure). Although the natural estimation strategy remains vulnerable to hazards after the problem is rid of degeneracies, our results suggest that it is still sensible even under adversarial conditions.

We still do not know the minimax strategy for on-line Gaussian density estimation with arbitrary covariances – a question first posed by Warmuth and Takimoto (2000b) – although our work sheds some light on the problem. While using arbitrary-covariance multivariate Gaussians is a step forward from simpler distributions like the fixed-covariance Gaussian and Bernoulli, it would also be interesting to consider on-line estimation with other statistical models, such as low-dimensional Gaussians or a mixture of Gaussians. Extending the work on on-line PCA (Warmuth and Kuzmin, 2006) may be one approach for the first.

Acknowledgements. We are grateful to the anonymous reviewers for their helpful suggestions, to the Los Alamos National Laboratory for supporting the second author with a graduate fellowship, and to the NSF for grant IIS-0347646.

References

Azoury, K., Warmuth, M.: Relative loss bounds for on-line density estimation with the exponential family of distributions. Journal of Machine Learning 43(3), 211–246 (2001)

Cesa-Bianchi, N., Lugosi, G.: Prediction, Learning, and Games. Cambridge University Press, Cambridge (2006)

Crammer, K.: Online tracking of linear subspaces. 19th Annual Conference on Learning Theory (2006)

Freund, Y.: Predicting a binary sequence almost as well as the optimal biased coin. 9th Annual Conference on Computational Learning Theory (1996)

Hannan, J.: Approximation to Bayes risk in repeated play. In: M. Dresher, A. Tucker, P. Wolfe (Eds.), Contributions to the Theory of Games, vol. III, pp. 97–139 (1957)

Hazan, E., Kalai, A., Kale, S., Agarwal, A.: Logarithmic regret algorithms for online convex optimization. 19th Annual Conference on Learning Theory (2006)

Kalai, A., Vempala, S.: Efficient algorithms for the online decision problem. 16th Annual Conference on Learning Theory (2005)

Shalev-Shwartz, S., Singer, Y.: Convex repeated games and Fenchel duality. Advances in Neural Information Processing Systems 19 (2006)

Takimoto, E., Warmuth, M.: The last-step minimax algorithm. 11th International Conference on Algorithmic Learning Theory (2000a)

Takimoto, E., Warmuth, M.: The minimax strategy for Gaussian density estimation. 13th Annual Conference on Computational Learning Theory (2000b)

Warmuth, M., Kuzmin, D.: Randomized PCA algorithms with regret bounds that are logarithmic in the dimension. Advances in Neural Information Processing Systems 19 (2006)

Zinkevich, M.: Online convex programming and generalized infinitesimal gradient ascent. In: 20th International Conference on Machine Learning (2003)

Appendix

Lemma 6. *For any* $n \in \mathbb{N}$,

$$(n+1)\ln\frac{n+1}{n} = 1 + \frac{1}{2}\ln\frac{n+1}{n} + s(n) - s(n+1)$$

where $s(n) = 1/(12n) - 1/(360n^3) + \ldots$ *is (the tail of) Stirling's series.*

Proof. Apply Stirling's formula: $\ln n! = n\ln n - n + (1/2)\ln(2\pi n) + s(n)$. □