Signal Intensity Deconvolution in Optical Receivers

Conner DiPaolo and Ryan Rogalin

***Abstract*—In a photon-counting free-space optical communication receiver, the signal energy should be centered on the photodetector in order to maximize the received signal power. Random spatial fluctuations of the beam can be induced by atmospheric as well as mechanical phenomena. In order to control a compensator for driving the beam to the center of the detector, we must be able to infer the location of the centroid with only rudimentary spatial information (the number of photons that have hit each *quadrant* of the detector). We close this control loop by making a reduction to the task of signal intensity reconstruction via an inverse problem, proving non-asymptotic estimation error bounds for a previously published method and novel simplification, and presenting two novel estimation schemes that empirically perform better in near- and medium-range deep space settings. Our centroid estimation method is the first procedure in the literature proved to be consistent.**

***Index Terms*—Optical communications, control, inverse problems, statistical estimation, sample complexity.**

I. Introduction

MUCH of deep space exploration has relied on radio transmission for communication, yet higher data requirements have pushed the communication model towards the better performing optical regime [1, Fig 1.1]. For a comprehensive survey, see [1]. For example, NASA's Lunar Laser Communications Demonstration in 2013 was able to achieve more than six times the best bit-rate of any radio link at similar distances [2]. For these reasons and more, NASA has decided to fly the Deep Space Optical Communication (DSOC) instrument on the upcoming 2022 Psyche mission at the unprecedented distance of 3AU [3].

Prior Art. Being able to focus, and as a necessity determine the location of, the photon beam of an optical communication link into the receiver is paramount, since extremely small spot sizes (compared to radio communications [1, Ch 5.3]) limit error tolerance. For quadrant separated receivers, Gagliardi and Sheikh [4] suggest simply differencing the number of incident photons above and below the x- and y-axes, respectively, to get rudimentary information about the location of the centroid. At far range with commensurately low signal-to-noise ratio this method has high variance [4, Fig. 4]. Analysis in [5] locally characterized the stability of *jointly* acting control systems relying on these estimates, but high variance in the input estimates can ensure that the transient, erroneous behavior of the system dominates the steady-state as seen in [4]. Other work in this area deals with high resolution receivers at close range, where the Gagliardi-Sheikh approach or a center of mass computation is employed. See, for example, [6] and [7, Fig 7] from the communications community and [8] from the optics community. Tracking Earth from a spacecraft has similar requirements; the standard method simply correlates the receiver photon image with a reference image of Earth [1, Ch 5.3.5.1.1] [9]. This is irrelevant in our case since the spacecraft is miniscule and occluded by atmosphere. Recent work of [10, 11] uses maximum likelihood centroid estimates to track the beam, but this work assumes a-priori knowledge of the background and signal intensity which in our case is not available; a *lower* bound on tracking convergence is presented but no upper bound on convergence is given.

Contributions. We frame centroid estimation (and hence, centroid control) as finding the proportion of signal photons in each receiver quadrant through an inverse problem. We show that consistent estimators of these proportions yield consistent estimators of the centroid, something that cannot be said for the Gagliardi-Sheikh centroiding method, or the numerical center-of-mass approach, which both retain intrinsic error from discretization, and in the latter case even forces estimates to be in $[-R, R] \times [-R, R]$ for a radius R detector [12, Fig 10.15] [6, Fig 8]. This hands us a natural reduction to the problem of deconvoluting signal intensity K_s from background intensity K_b in each quadrant. In this realm, we derive non-asymptotic sample complexity and $\mathcal{L}_2$ risk bounds for a previously published method and a novel simplification, which hand practitioners straightforward guidance on integration times needed to estimate K_b and K_s. We also propose two non-consistent methods for K_b and K_s estimation that empirically outperform the former in low and medium range settings at low integration times. Guarantees help decoding too, since this relies on K_b, K_s estimates through both channel estimation and the decoding algorithm [13].

Notation. $\mathbf{0}_n, \mathbf{1}_n \in \mathbb{R}^n$ are the all-zero and -ones vectors. $\oplus$ gives the concatenation of two vectors.

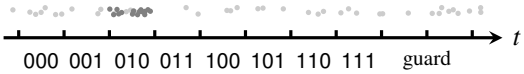
II. The Inverse Problem

In the typical Pulse Position Modulation optical communications model [12, Ch 6.7], pulses of photons are sent during one of M time slots, each corresponding to a specific bit sequence for the symbol. With the addition of Inter-Symbol Guard Times, $P < M$ time slots that never receive a photon pulse are appended to each symbol. This is done for statistical timing synchronization [14]. In the example symbol below, $M = 8$ and $P = 2$, while the signal slot $j = 3$ corresponding to the bits 010 is transmitted.

This publication was prepared by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration and a California Institute of Technology SURF Fellowship. Copyright 2019 California Institute of Technology. U.S. Government sponsorship acknowledged.

Conner DiPaolo was with the Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA during the creation of this work. He is now with the Department of Mathematics, Harvey Mudd College, 301 Platt Boulevard, Claremont, CA. (email: cdipaolo@hmc.edu)

Ryan Rogalin is with the Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA. (email: Ryan.Rogalin@jpl.caltech.edu)



The background and signal photon intensities in each slot are K_b and K_s, respectively. Under the Poisson model [1, 12], the number of photons received in slot $i \in \{1, \ldots, M+P\}$ are distributed $X_i \sim \text{Pois}(K_b + \delta_{ij} K_s)$, conditioned on the uniformly distributed signal slot $j \sim \text{Unif}\{1, \ldots, M\}$. We will largely ignore the estimation of the timing offset of the first slot.

Write D for the radius R circle centered at the origin and D_i for the i-th quadrant of D. The proportion of photons from the communications beam hitting each quadrant $i \in \{1, 2, 3, 4\}$ in the receiver D is

$$P_i(x_0, y_0) = \iint_{D_i} \phi\left(\tfrac{x-x_0}{\sigma}\right)\phi\left(\tfrac{y-y_0}{\sigma}\right) dA \left(\iint_{D} \phi\left(\tfrac{x-x_0}{\sigma}\right)\phi\left(\tfrac{y-y_0}{\sigma}\right) dA \right)^{-1}$$

collected into a vector $\boldsymbol{p} = (P_1, \ldots, P_4)$, where ϕ is the zero mean unit variance Gaussian density function, $(x_0, y_0) \in \mathbb{R}^2$ is the beam center relative to the receiver, and σ is the beam width. This can be computed numerically. In the case with noisy observations $\widetilde{\boldsymbol{p}}$, a simple and fast way to find the generating centroid is to solve a least-squares problem

$$(\hat{x}_0, \hat{y}_0) = \underset{(x_0, y_0)}{\arg\min} \left\| \boldsymbol{p}(x_0, y_0) - \widetilde{\boldsymbol{p}} \right\|_2^2 = \underset{(x_0, y_0)}{\arg\min}\, f(x_0, y_0). \quad (1)$$

Empirically, f is quasi-convex within a large berth around the receiver. Take $\widetilde{\boldsymbol{p}} = \boldsymbol{p}(\frac{1}{2}, \frac{1}{2})$, $R = 2\sigma = 1$, and vary $f(x_0, y_0)$ along $x_0 = y_0$ to exhibit non-quasi-convexity at around $x_0 = -2R$. To make consistency guarantees about $(\hat{x}_0, \hat{y}_0)$ we need to know that $\boldsymbol{p} : \mathbb{R}^2 \to \mathbb{R}^4$ is injective. This global identification is a surprisingly difficult verification as often seen in econometrics in the context of the Generalized Method of Moments [15, p.2127]. Luckily, we can make an analytical argument (in the Appendix) that gives this result.

Lemma 1. $\boldsymbol{p} : \mathbb{R}^2 \to \mathbb{R}^4$ is injective.

Theorem 2. If $\widetilde{\boldsymbol{p}} \to \boldsymbol{p}$ in measure, in distribution, or almost surely as the number of observed symbols increases, then $(\hat{x}_0, \hat{y}_0) \to (x_0, y_0)$ in the same manner.

Proof. $\|\boldsymbol{p}(x_0', y_0') - \boldsymbol{p}(x_0, y_0)\|_2^2$ is minimized only when $(x_0', y_0') = (x_0, y_0)$ by Lemma 1. Joint continuity of f ensures that $(\hat{x}_0, \hat{y}_0)$ is continuous in $\widetilde{\boldsymbol{p}}$ at $\widetilde{\boldsymbol{p}} = \boldsymbol{p}$ by Corollary 8.1 of [16]. Applying the Mann-Wald Theorem [17] guarantees that this continuous function $(\hat{x}_0, \hat{y}_0)$ of a convergent random variable $\widetilde{\boldsymbol{p}} \to \boldsymbol{p}$ converges to (x_0, y_0) in measure, distribution, or almost surely so long as $\widetilde{\boldsymbol{p}} \to \boldsymbol{p}$ with the same mode of convergence. ∎

Since consistent estimates $\widetilde{K}_i$ of K_s in each D_i grant consistent estimates of $\boldsymbol{p}$ as $\widetilde{\boldsymbol{p}}_i = \widetilde{K}_i / (\widetilde{K}_1 + \cdots + \widetilde{K}_4)$ by the same Mann-Wald Theorem [17], Theorem 2 suggests that improved algorithms for K_s estimation can result in more efficient centroid estimates. This follows since accelerated convergence of K_s would directly accelerate convergence of the estimated slot counts $\widetilde{\boldsymbol{p}}_i$, which carries through to our estimates of (x_0, y_0) by Theorem 2.

III. K_s Estimation and Sample Complexity

Like most work on guaranteed estimation of inverse problems, estimation of K_b and K_s relies on structural assumptions about the signal. Luckily, strong structural guarantees given by the inter-symbol guard time slots make coming up with viable estimators relatively straightforward. Bounds, on the other hand, take much more work. Let $\boldsymbol{y} \in \mathbb{N}^{M+P}$ be a vector corresponding to the total photon count in each time slot over n symbols. The elements Y_i of $\boldsymbol{y}$ are distributed according to the following generative process:

(1) Draw pulse count $\boldsymbol{n} \sim \text{Multinomial}(n, \frac{1}{M}\mathbf{1}_M \oplus \mathbf{0}_P)$,
(2) Draw photon counts $Y_i \sim \text{Pois}(nK_b + \boldsymbol{n}_i K_s)$.

Maximum likelihood or Bayesian estimation of K_b and K_s (see, Section IV) is intractable unless we know or can accurately estimate the bit sequence. In the presence of high background radiation this gives high bias, so in practice estimators based on arguments using the law of large numbers are used. For example, the law of large numbers guarantees $\frac{1}{n}Y_i \to K_b + \frac{1}{M}K_s$ for $i = 1, \ldots, M$ and $\frac{1}{n}Y_i \to K_b$ for $i = M+1, \ldots, M+P$ almost surely as $n \to \infty$, so the following estimators for K_b and K_s are strongly consistent:

$$\hat{K}_b = \min_j \frac{1}{n(P-k)} \sum_{i=1}^{P-k} Y_{i+j \bmod M+P},$$

$$\hat{K}_s = \max_j \frac{M}{n(M-k)} \sum_{i=1}^{M-k} Y_{i+j \bmod M+P} - M\hat{K}_b.$$

The maximum is done to account for gross timing offset errors in practice, and $k \in \{0, 1\}$ is set to $k = 1$ to account for fractional errors in the timing offset. These estimators were written with the incorrect scaling in [13]; The above formulation corrects that typographic mistake. We can retain the performance (both empirically and in terms of strong consistency) in a simpler version of the above scheme by instead using $\widetilde{K}_s = \frac{1}{n}\sum_{i=1}^{M+P} Y_i - (M+P)\hat{K}_b$. This will be called the improved convolutional scheme, differentiated symbolically with $\widetilde{K}_s$ instead of $\hat{K}_s$. It non-trivially faster to compute, requiring one fewer convolution.

For analysis, we start with some lemmata, the first of which may be of independent interest. The following proofs rely heavily on bounding the Cramér transforms of random variables. For more details on this technique, see [18] or the proofs in the Appendix.

Lemma 3. Let $X_i \sim \text{Pois}(\lambda_i)$ for $i = 1, \ldots, N$ where $\lambda_{\min} = \min \lambda_i \geq \log(N) + 1$ and $\lambda_{\max} = \max \lambda_i > 2 \log N$. Then

$$\mathbb{E} \min X_i \geq \lambda_{\min} - \log(N+1), \quad \mathbb{E} \max X_i \leq \frac{\lambda_{\max} - \log N}{1 - \sqrt{\frac{2 \log N}{\lambda_{\max}}}}.$$

Lemma 4. Let $\boldsymbol{n} \sim \text{Multinomial}(n, \frac{1}{M}\mathbf{1}_M \oplus \mathbf{0}_P)$ and $\boldsymbol{n}_*^{(j)} = \sum_{i=0}^{M-k-1} \boldsymbol{n}_{i+j \bmod M+P}$ for all $j = 0, 1, \ldots, M+P-1$ and $k \in \{0, 1\}$. If $\boldsymbol{m}_* = \frac{1}{n(M-k)} \max \boldsymbol{n}_*^{(j)} - \frac{1}{M}$ then

$$\mathbb{E}\boldsymbol{m}_* < \frac{k}{M-k}\sqrt{\frac{2}{n}} + \frac{k}{\sqrt{2n}}(M-P)e^{-\frac{nP^2}{2M^2}}.$$

Moreover, if $n \geq 2\frac{M^2}{P^2}\log\left(\frac{M(M+P)}{2}\right)$ then $\mathbb{E}\boldsymbol{m}_* < \frac{k}{M-k}\sqrt{\frac{8}{n}}$.

The following Corollary combines Jensen's inequality applied to $f(x_1, \ldots, x_n) = \max\{x_1, \ldots, x_n\}$ with the above Lemmas to bound the bias of $\hat{K}_b$, $\hat{K}_s$, and $\widetilde{K}_s$.

Corollary 5. Write $S = M + P$ for clarity.

$$\max\left\{|\mathbb{E}\hat{K}_b - K_b|, \frac{|\mathbb{E}\widetilde{K}_s - K_s|}{S}\right\} \leq \frac{\log(S+1)}{n(P-k)}$$

whenever $n \geq \frac{\log(S)+1}{(P-k)K_b}$, and

$$|\mathbb{E}\hat{K}_s - K_s| \leq \frac{M\log(S+1)}{n(P-k)} + \frac{kK_s}{M-k}\sqrt{\frac{32}{n}} + \frac{MK_b+K_s}{\sqrt{n}}\sqrt{\frac{8\log S}{(M-k)K_b+K_s}}$$

when $n \geq \max\left\{\frac{\log(S)+1}{(P-k)K_b}, \frac{8\log S}{(M-k)K_b+K_s}, 2\frac{M^2}{P^2}\log\left(\frac{MS}{2}\right)\right\}$.

Bounding the variance of these estimators is much simpler through use of the Efron-Stein inequality [18, Thm 3.1].

Lemma 6. Denote $K = K_s + K_s^2$. Then

$$\max\left\{\mathrm{Var}(\hat{K}_b), \frac{\mathrm{Var}(\widetilde{K}_s)}{(M+k)^2}, \frac{\mathrm{Var}(\hat{K}_s)}{4M^2}\right\} \leq \frac{(M+P)K_b + K}{n(P-k)^2}.$$

Proof. If X and X' are i.i.d then $\mathbb{E}(X - X')^2 = 2\,\mathrm{Var}(X)$. If we only consider one slot y_i, the functional

$$\hat{K}_b(\ldots, y_i, \ldots) = \min_j \frac{1}{n(P-k)} \sum_{i=1}^{P-k} y_{i+j \bmod M+P}$$

is a minimum of $n^{-1}(P-k)^{-1}$-Lipschitz functionals, and hence $n^{-1}(P-k)^{-1}$-Lipschitz. Write $\boldsymbol{y}_i'$ for $\boldsymbol{y}$ with the i-th coordinate Y_i replaced with an independent copy. The Lipschitz property, and the Efron-Stein inequality guarantee

$$\mathrm{Var}(\hat{K}_b) \leq \frac{1}{2}\sum_{i=1}^{M+P} \mathbb{E}[(\hat{K}_b(\boldsymbol{y}) - \hat{K}_b(\boldsymbol{y}_i'))^2] \leq \frac{\sum_{i=1}^{M+P}\mathrm{Var}(Y_i)}{n^2(P-k)^2}$$

Calculating $\mathrm{Var}(Y_i) = nK_b + \frac{n}{M}K_s + \frac{n}{M}(1 - \frac{1}{M})K_s^2$ for $i = 1, \ldots, M$ and $\mathrm{Var}(Y_i) = nK_b$ otherwise gives us the result. The arguments for $\widetilde{K}_s$ and $\hat{K}_s$ are similar. ∎

The bias-variance decomposition of $\mathcal{L}_2$ loss gives concentration by a Chebyshev-style bound. We note that one can directly prove exponential concentration bounds for $\hat{K}_b$, $\hat{K}_s$, and $\widetilde{K}_s$ with $O(\log \delta^{-1})$ sample-complexity via sub-exponentiality of the Poisson and Binomial, but a seemingly necessary triangle inequality application makes this unbearably loose for practical parameter settings.

Theorem 7. Denote $K = K_s + K_s^2$. The mean squared error

$$\mathbb{E}(\widetilde{K}_s - K_s)^2 \leq \frac{2(M+P)^2}{n(P-k)^2}\left[(M+P)K_b + K\right]$$

whenever $n \geq \max\left\{\frac{\log(M+P)+1}{(P-k)K_b}, \frac{\log^2(M+P+1)}{(M+P)K_b+K}\right\}$. In particular, ensuring n is in this valid range and

$$n \geq \frac{2(M+P)^2}{\epsilon^2\delta(P-k)^2}\left[(M+P)K_b + K\right]$$

ensures $|\widetilde{K}_s - K_s| < \epsilon$ with probability at least $1 - \delta$.

For $\hat{K}_b$ and $\hat{K}_s$, adding the square of the bias from Corollary 5 and variance from Lemma 6 gives us similar $\mathcal{L}_2$ risk bounds and concentration results. Known scenario parameters (see Section V) can be used to turn these bounds into guidance.

IV. Faster K_s Estimation at Medium Range

While the estimation scheme presented in Section III is guaranteed to work with the asymptotically minimax rate for Poisson mean estimation, reliance of the convergence of the slot counts can be improved in the case that we know (or can estimate easily as $\hat{j} = \arg\max X_i$) the signal slot. The simplest setup is to include j during maximum likelihood estimation of $(\{j^{(n)}\}, K_b, K_s)$ given symbols $(X_i^{(n)})$. One can show that this estimator satisfies the online updates

$$\hat{K}_b^{(n)} = \frac{n-1}{n}\hat{K}_b^{(n-1)} + \frac{1}{n(M+P-1)}\left[\sum_{i=1}^{M+P} X_i^{(n)} - X_{\hat{j}^{(n)}}^{(n)}\right],$$

$$\hat{K}_s^{(n)} = \frac{n-1}{n}\left[\hat{K}_s^{(n-1)} + \hat{K}_b^{(n-1)}\right] + \frac{1}{n}X_{\hat{j}^{(n)}}^{(n)} - \hat{K}_b^{(n)},$$

via a standard calculus-based argument (as in [19, Ch 9.3]) using the stationarity of K_b and K_s. By including the signal slots $\{j^{(n)}\}$ in the maximum likelihood computation we implicitly assume that the slot which receives the most photons is the signal slot. If $\hat{j}^{(n)} = j^{(n)}$, the estimators are unbiased and their mean squared errors are both at most $n^{-1}(K_b + K_s)$. In reality, $\hat{j}$ is a poor estimate when K_b/K_s is large; $\hat{K}_s^{(n)}$ over-estimates K_s and $\hat{K}_b^{(n)}$ under-estimates K_b in expectation.

An Approximate Variational Filter (AVF) is another possible solution. Here, at each symbol we update our beliefs about K_b and K_s by assuming $j = \hat{j}$. This is intractable to maintain explicitly, but repeated projection of posteriors into the prior space maintains tractability. This projection step is called variational inference; for the projection we simply moment match [20]. For our model, if we posit independent Gamma priors $K_b \sim \mathrm{Gamma}(\alpha_b, \beta_b)$ and $K_s \sim \mathrm{Gamma}(\alpha_s, \beta_s)$ we can expand the posterior as a mixture of Gamma distributions by classically applying Bayes' rule and using the binomial theorem to reduce the signal slot term $(K_b + K_s)^{X_j}$ into powers of K_b and K_s. The resulting posterior is generated by the following process:

(i) Draw $i \sim \mathrm{Categorical}(w_k)$,
(ii) Draw $K_b|i, j, \{X_k\} \sim \mathrm{Gamma}(X_{-j} + \alpha_B + i, M + P + \beta_b)$ and $K_s|i, j, \{X_k\} \sim \mathrm{Gamma}(X_j + \alpha_b - i, 1 + \beta_s)$.

where the mixing weights

$$w_k \propto \binom{X_j}{k}\frac{\Gamma(X_{-j} + \alpha_b + i)\Gamma(X_j + \alpha_s - i)}{(M + P + \beta_b)^{X_{-j}+\alpha_b+i}(1+\beta_s)^{X_j+\alpha_s-i}}$$

and $X_{-j} = \sum_{k\neq j} X_k$. From the above we can readily compute the posterior mean and variances in order to moment match back to Gamma distributions. In the posterior, i corresponds to how many photons in the signal slot belong to the background noise. Consideration of this trade-off allows the AVF to slightly outperform the maximum likelihood estimator (MLE).

V. Empirical Performance

We evaluate the performance of the algorithms in three regimes of interest for deep space: At near-Earth range, $K_b \approx 0.00239$, $K_s \approx 2.56$, and $M = 16$. At medium range, $K_b \approx 0.00869$, $K_s \approx 2.83$, and $M = 128$. At far range, $K_b \approx 0.04575$, $K_s \approx 0.395$, and $M = 128$. For regulatory reasons, only slot widths with $M = 4P$ are used, so we only show these parameter settings in our computational examples.

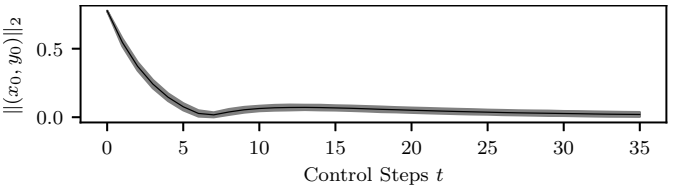


Fig. 1. Sample average $\pm 2\hat{\sigma}$ of centroid norm of an idealized receiver undergoing PID control with $K_P = 0.2$, $K_I = 0.1$, and $K_d = 0$, and centroid starting from $(x_0, y_0) = (-\frac{3}{4}, \frac{1}{5})$ at medium range. The receiver $R = 2\sigma = 1$. K_s estimation is done via the $\hat{K}_s$ scheme over $n = 1000$ symbols.

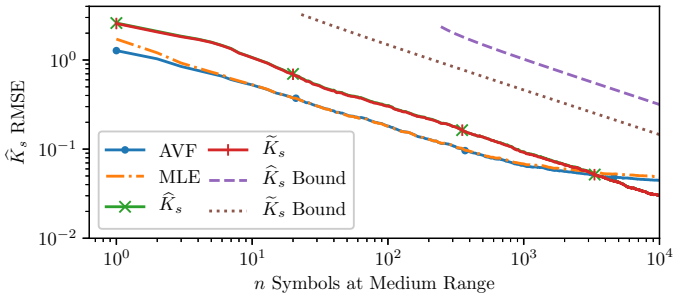


Fig. 2. Medium range comparison of different estimation algorithms for K_s. Note that the AVF and MLE estimation schemes outperform the convolutional estimation for $n \lesssim 2000$ before slot estimation bias dominates.

With noise, centroid estimates tend to shrink towards the origin, but the angular bias remains quite small ($\sim 10^{-2}$ radians if we assume K_s estimates are distributed as truncated normals with the correct proportion mean and standard deviation 0.3). As a result, simulating a control environment (Figure 1) in the medium range regime returns fast convergence that is highly robust to the noise in the K_s estimates, even though (Figure 2) the relative error in the K_s estimates at the chosen $n = 1000$ integration time is on the order of 5-10%. The increase in error around step $t = 7$ corresponds to overshooting the origin, and could be reduced by tuning the control parameters further.

In terms of Root Mean Squared Error (RMSE), the medium range scenario benefits most from the acceleration described in Section IV (Figure 2). At close range, the Section IV estimators consistently outperform the convolutional estimation of Section III by a multiplicative factor of about 1.2. In deep space, where background intensity starts to dominate, the convolutional estimators outperform the accelerated scheme as early as $n = 100$ symbols. For K_s estimation, the upper bound on the $\widetilde{K}_s$ performance beats the Section IV estimators at $n \sim 200$. For $\hat{K}_s$ this happens at $n \sim 1000$ symbols. The upper bounds on performance derived in Section III get quite tight as we move to far range, where the bound on $\widetilde{K}_s$ tracks the empirical value within a multiplicative factor of about 1.7.

VI. Conclusion

We have to the best of our knowledge introduced the first provably consistent algorithm for optical communications centroiding when signal and background intensities are unknown a-priori, and provided explicit rates of convergence for the subproblem of K_s estimation in PPM links, as well as accelerated algorithms that perform better at up to medium range. Since decoding relies on K_b and K_s estimation to create log-likelihood ratios and timing estimates [13], our guarantees also help these related endeavors. Using our improved estimators of the signal and background intensities could also make the work of [10, 11] more practical through not relying on a-priori knowledge of these constants, but this and creating convergence guarantees of the resulting centroid estimator is left for future work.

Appendix: Additional Proofs

Proof of Lemma 1. Without loss of generality take $\sigma = 1$. Define $I_i(x_0, y_0) = \iint_{D_i} \phi(x-x_0)\phi(y-y_0)\,dx\,dy$ and $I(x_0, y_0) = \iint_D \phi(x-x_0)\phi(y-y_0)\,dx\,dy$. By the quotient rule we know that $\partial_{x_0} P_i(x_0, y_0) > 0$ if and only if $I\partial_{x_0} I_i - I_i \partial_{x_0} I > 0$, or equivalently if $\frac{\partial_{x_0} I_i}{I_i} > \frac{\partial_{x_0} I}{I}$. But since the common integrand defining I_i and I is uniformly positive we know $I_i < I$, or $1/I < 1/I_i$, and hence $\partial_{x_0} P_i(x_0, y_0) > 0$ if

$$\partial_{x_0} I - \partial_{x_0} I_i = \iint_{D\setminus D_i} (x - x_0)\,\phi(x-x_0)\phi(y-y_0)\,dx\,dy$$

is strictly negative. Consider now $i = 1$ and $x_0 > 0$. For any $x > 0$ and arbitrary y, $\phi(x-x_0)\phi(y-y_0) \geq \phi(-x-x_0)\phi(y-y_0)$. Thus Fubini's Theorem tells us

$$\partial_{x_0} I - \partial_{x_0} I_i \leq \int_{-R}^{R}\int_{\mathbb{R}} (x - x_0)\phi(x-x_0)\phi(y-y_0)\,dx\,dy$$
$$= \int_{-R}^{R} (x - x_0)\phi(x-x_0)\,dx = \tfrac{-1}{\sqrt{2\pi}} e^{-\frac{1}{2}(R+x_0)^2}\left(e^{2Rx_0} - 1\right),$$

which is strictly negative. To summarize, this combined with symmetry ensures $\partial_{x_0} P_1 > 0$ and $\partial_{y_0} P_1 > 0$ uniformly. Similarly, $\partial_{x_0} P_4 > 0$ and $\partial_{y_0} P_4 < 0$ uniformly.

Now we prove injectivity. Our derivative information and symmetry guarantee that $P_i(x_0, y_0) = \max_j P_j(x_0, y_0)$ if and only if (x_0, y_0) are in the i-th orthant, so take $(x_0, y_0) \neq (x_0', y_0')$ in the first orthant and define $f_i(t) = P_i(tx_0' + (1-t)x_0, ty_0' + (1-t)y_0)$ for all $t \in \mathbb{R}$. Without loss of generality, two scenarios are possible: $x_0 \leq x_0'$ and $y_0 \leq y_0'$, or $x_0 \leq x_0'$ and $y_0 \geq y_0'$. In the former case, $f_1'(t) > 0$ for all $0 < t < 1$, and hence $P_1(x_0, y_0) = f_1(1) \neq f_1(0) = P_1(x_0', y_0')$. In the latter, $f_4'(t) > 0$ for $0 < t < 1$ and similarly $P_4(x_0, y_0) = f_4(1) \neq f_4(0) = P_1(x_0', y_0')$. In any case, $\boldsymbol{p}(x_0, y_0) \neq \boldsymbol{p}(x_0', y_0')$. ∎

Proof of Lemma 3. From the moment generating function of a Poisson random variable [19] we can compute

$$\varphi_{\lambda_{\min} - X_i}(t) = \log \mathbb{E} e^{t(\lambda_{\min} - X_i)} \leq t\lambda_{\min} + \lambda_{\min}(e^{-t} - 1) =: \varphi(t)$$

To compute the Fenchel-Legendre conjugate of this φ we just need to perform a simple univariate optimization, verifiable by a computer algebra system: $\varphi^*(t) = \sup_{\alpha \geq 0}\left(\alpha t - \varphi(\alpha)\right) = \log\left(\frac{\lambda_{\min}}{\lambda_{\min} - t}\right) + t$ defined on $t \in [0, \lambda_{\min})$. This φ^* has a functional inverse $(\varphi^*)^{-1}(\log N) = \lambda_{\min} - W\left(\frac{e^{\lambda_{\min}}\lambda_{\min}}{N}\right)$ where $W(z)$ is the Lambert-W function (see [21]). Theorem 2.5 of [18] then ensures

$$\mathbb{E}\min_{1\leq i\leq N} X_i = \lambda_{\min} - \mathbb{E}\max_{1\leq i\leq N}(\lambda_{\min} - X_i) \geq \lambda_{\min} - (\phi^*)^{-1}(\log N)$$
$$= W\left(\tfrac{e^{\lambda_{\min}}\lambda_{\min}}{N}\right) \geq \lambda_{\min} + \log\frac{\lambda_{\min}}{N} - \log\left(\lambda_{\min} + \log\frac{\lambda_{\min}}{N}\right)$$
$$\geq \lambda_{\min} - \log(N+1).$$

The bound of the Lambert-W function comes from [21] and holds so long as $\lambda_{\min} \geq \log(N) + 1 > W(eN)$.

For the expected maximum we can play a similar game. In particular, consider $\varphi_{X_i}(t) = \log \mathbb{E} \exp\big(tX_i\big) \leq \lambda_{\max}(e^t - 1) =: \gamma(t)$. This function γ has Fenchel-Legendre conjugate

$$\gamma^*(t) = \sup_{\alpha \geq 0} \big(\alpha t - \gamma(\alpha)\big) = \lambda_{\max} - t + t \log \frac{t}{\lambda_{\max}}$$

and respective inverse

$$(\gamma^*)^{-1}(y) = \frac{y - \lambda_{\max}}{W\big(\frac{y-\lambda_{\max}}{e\lambda_{\max}}\big)},$$

both computable by simple calculus. The scalar inequality $-1/W(x) \leq 1/(1 - \sqrt{2}\sqrt{1 + ex})$ for $-e^{-1} \leq x < -\frac{1}{2}e^{-1}$ can be verified in a computer algebra system, which applied to $(\gamma^*)^{-1}$ ensures $(\gamma^*)^{-1}(y) \leq \big(1 - \sqrt{2}\sqrt{\frac{y}{\lambda_{\max}}}\big)^{-1}\big(\lambda_{\max} - y\big)$ so long as $\lambda_{\max} > 2y$. Theorem 2.5 of [18] again guarantees

$$\mathbb{E} \max_{1 \leq i \leq N} X_i \leq (\gamma^*)^{-1}(\log N) \leq \frac{\lambda_{\max} - \log N}{1 - \sqrt{2}\sqrt{\frac{\log N}{\lambda_{\max}}}}$$

under our condition that $\lambda_{\max} > 2 \log N$. ∎

Proof of Lemma 4. The situation when $k = 0$ is clear since the variables $\boldsymbol{n}_*^{(j)}$ are equal in which case $\mathbb{E}\boldsymbol{m}_* = 0$, so consider the other case $k = 1$. By Hoeffding's inequality [18, Sec 2.3] applied to the binomial random variables $n_*^{(j)} \sim \text{Binomial}(n, \frac{1}{M}|\{i : i = 1, 2, \ldots, M - k,\ 0 \leq i + j \bmod M + P \leq M - 1\}|)$ we know

$$\begin{aligned}
\mathbb{P}\big(\tfrac{1}{n(M-k)}\boldsymbol{n}_*^{(j)} \geq \tfrac{1}{M} + t\big) &= \mathbb{P}\big(\tfrac{1}{n}\boldsymbol{n}_*^{(j)} \geq \tfrac{M-k}{M} + (M-k)t\big) \\
&\leq \mathbb{P}\big(\tfrac{1}{n}\boldsymbol{n}_*^{(j)} \geq \tfrac{1}{n}\mathbb{E}\boldsymbol{n}_*^{(j)} + (M-k)t\big) \\
&\leq e^{-2n(M-k)^2t^2}.
\end{aligned}$$

Since $P < M$, $2P$ of the $\boldsymbol{n}_*^{(j)}$ variables contain summands from components of $\boldsymbol{n}$ that are zero almost surely and hence are always less than neighboring windows. There are then at most $M + P - (2P) = M - P$ slots which can have the maximum, so $\mathbb{P}\big(\boldsymbol{m}_* \geq t\big) \leq (M - P)e^{-2n(M-k)^2t^2}$ by a union bound. Integrating gives

$$\begin{aligned}
\mathbb{E}\boldsymbol{m}_* &= \int_0^\infty \mathbb{P}\big(\boldsymbol{m}_* \geq t\big)\, dt \leq (M - P) \int_0^\infty e^{-2n(M-k)^2t^2}\, dt \\
&= k\frac{M - P}{M - k}\sqrt{\frac{\pi}{8n}} < \frac{k}{\sqrt{2n}}.
\end{aligned}$$

But the only time the maximal $\boldsymbol{n}_*^{(j)}$ won't occur at $j \in \{0, 1\}$ is when one of the end signal slots, say $\boldsymbol{n}_1$, is larger than the sum, say $N = \boldsymbol{n}_2 + \cdots + \boldsymbol{n}_{P+2}$ of $P + 1$ signal slots. Since $N - \boldsymbol{n}_1 \in [-n, n]$ and $\mathbb{E}N - \boldsymbol{n}_1 = \frac{P}{M}$, Hoeffding tells us again that this happens with probability at most $(M-P)\exp\big(-\frac{nP^2}{2M^2}\big)$. When the maximum occurs in the first two slots, $\mathbb{E}[\boldsymbol{m}_*|j_* \in \{0, 1\}] \leq \frac{k}{M-k}\sqrt{\frac{\pi}{2n}} < \frac{k}{M-k}\sqrt{\frac{2}{n}}$. The final bound comes from conditioning on this event in computing the expectation. ∎

Acknowledgement

The authors would like to thank Zaid Towfic and Meera Srinivasan for helpful discussions during the process of this research.

References

[1] H. Hemmati, *Deep space optical communications*. John Wiley & Sons, 2006, vol. 11.
[2] L. David, "Historic demonstration proves laser communication possible," October 2013. [Online]. Available: https://www.nasa.gov/content/goddard/historic-demonstration-proves-laser-communication-possible
[3] ——, "Deep space communications via faraway photons," October 2017. [Online]. Available: https://www.jpl.nasa.gov/news/news.php?feature=6967
[4] R. Gagliardi and M. Sheikh, "Pointing error statistics in optical beam tracking," *IEEE Transactions on Aerospace and Electronic Systems*, no. 5, pp. 674–682, 1980.
[5] G. Marola, D. Santerini, and G. Prati, "Stability analysis of direct-detection cooperative optical beam tracking," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 25, no. 3, pp. 325–334, 1989.
[6] S. Lee, J. W. Alexander, and M. Jeganathan, "Pointing and tracking subsystem design for optical communications link between the international space station and ground," in *Free-Space Laser Communication Technologies XII*, vol. 3932. International Society for Optics and Photonics, 2000, pp. 150–158.
[7] D. Russell, H. Ansari, and C.-C. Chen, "Lasercom pointing, acquisition, and tracking control using a ccd-based tracker," in *Free-Space Laser Communication Technologies VI*, vol. 2123. International Society for Optics and Photonics, 1994, pp. 294–304.
[8] M. G. Spencer, B. N. Agrawal, M. Romano, R. L. Brunson, J. D. Dillow, D. H. Nelson, J. Connors, and S. R. Restaino, "Acquisition, tracking, pointing, and line-of-sight control laboratory experiments for a space-based bifocal relay mirror," in *Acquisition, Tracking, and Pointing XVI*, vol. 4714. International Society for Optics and Photonics, 2002, pp. 54–65.
[9] T.-Y. Yan, "Extended-source spatial acquisition process based on maximum-likelihood criterion for planetary optical communications," in *Free-Space Laser Communication Technologies IV*, vol. 1635. International Society for Optics and Photonics, 1992, pp. 273–286.
[10] M. S. Bashir and M. R. Bell, "Optical beam position estimation in free-space optical communication," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 52, no. 6, pp. 2896–2905, 2016.
[11] ——, "Optical beam position tracking in free-space optical communication systems," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 54, no. 2, pp. 520–536, 2018.
[12] R. M. Gagliardi and S. Karp, *Optical communications*. New York: Wiley-Interscience, 1976.
[13] M. Srinivasan, R. Rogalin, N. Lay, M. Shaw, and A. Tkacenko, "Downlink receiver algorithms for deep space optical communications," in *Proc. of SPIE Vol*, vol. 10096, 2017, pp. 100 960A–1.
[14] R. Rogalin and M. Srinivasan, "Synchronization for optical ppm with inter-symbol guard times," *Interplanetary Network Progress Report*, vol. 209, pp. 1–19, 2017.
[15] W. K. Newey and D. McFadden, "Large sample estimation and hypothesis testing," *Handbook of econometrics*, vol. 4, pp. 2111–2245, 1994.
[16] W. W. Hogan, "Point-to-set maps in mathematical programming," *SIAM review*, vol. 15, no. 3, pp. 591–603, 1973.
[17] H. B. Mann and A. Wald, "On stochastic limit and order relationships," *The Annals of Mathematical Statistics*, vol. 14, no. 3, pp. 217–226, 1943.
[18] S. Boucheron, G. Lugosi, and P. Massart, *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press, 2013.
[19] L. Wasserman, *All of statistics: a concise course in statistical inference*. Springer Science & Business Media, 2013.
[20] D. M. Blei, A. Kucukelbir, and J. D. McAuliffe, "Variational inference: A review for statisticians," *Journal of the American Statistical Association*, vol. 112, no. 518, pp. 859–877, 2017.
[21] A. Hoorfar and M. Hassani, "Approximation of the lambert w function and hyperpower function," *Research report collection*, vol. 10, no. 2, 2007.