Fast Gaussian Process Regression for Multiuser Detection in DS-CDMA

Dinh-Mao Bui, Sungyoung Lee

Abstract—Recently, Gaussian process has been excellently proved the effectiveness in solving the multiuser detection issue in Code Division Multiple Access system. Even with limited training sequences, Gaussian process-based solutions still surpass other approaches. However, due to the high complexity in terms of computation, the performance of this approach might be degraded. In this letter, we would like to propose an efficient method to reduce the complexity while still maintaining the desired accuracy. Finally, the proposed method is validated *via* the experiment.

Index Terms—Gaussian process regression, complexity reduction, multiuser detection, DS-CDMA, hyper-parameters estimation, signal interference.

1 Introduction

The direct sequence code division multiple access (DS-CDMA) system distinguishes users by signals over the channel. Unfortunately, the interference between the signals happens even with a small number of users and can be recognized as the multiple access interference (MAI). This noisy issue critically increases the bit error rate (BER) under the near/far effect. To alleviate this issue, the multiuser detection [1] (MUD) technique has been developed to reduce the interference. The known optimal solution for MUD can be retrieved *via* minimizing the mean square error (MMSE) estimation [2]. Nevertheless, doing this estimation immensely costs the computational resources. In order to solve this problem, many approaches have been proposed. Among these approaches, Gaussian process regression (GPR) is considered as the most promising method in terms of flexibility and accuracy [3].

Practically, GPR is widely used in many research fields such as data communication, networking and signal processing. Rather than finding exact parameters for the model, GPR helps to adapt the parameters to represent the underlying function. As such, GPR is a suitable choice for noisy, corrupted or erroneous input data. However, this method encounters a critical drawback, namely the high complexity. Theoretically, GPR requires $O(n^3)$ for computation when calculating n training points of dataset. In this research, we propose a method to reduce the complexity to $O(n \log n)$. Consequently, this improvement significantly accelerates the regression while still maintaining the analogous BER compared with the MMSE estimation.

2 Regression Model

2.1 Assumption

The target system is a synchronous DS-CDMA system. This system is designed to serve a number of users on the same channel. To differentiate the signals, the spreading codes are assigned to the users. Initially, these codes are multiplied with the up-sampled original signals. Subsequently, the chips for these signals are transmitted through the channel. Since the channel is linear and noisy, the separated chips are combined together including a known additive white Gaussian noise (AWGN). In the end, the MUD takes responsibility to recover the original signals from the received chips.

Assume that the input dataset with n training points is denoted by $\mathcal{D} \equiv \{\boldsymbol{x}_t, \boldsymbol{y}_t|_{t=1,\dots,n}\}$, where $\boldsymbol{x}_t$ is the vector of original signals, $\boldsymbol{y}_t$ is the vector of received chips. Both $\boldsymbol{x}_t$ and $\boldsymbol{y}_t$ are collected at time t. The relationship of these vectors can be shown in matrix notation as follows:

$$\boldsymbol{y}_t = \boldsymbol{S}\,\boldsymbol{A}\,\boldsymbol{x}_t + \boldsymbol{n}_t, \tag{1}$$

where $\boldsymbol{S}$ is an $U \times V$ matrix (each column of this matrix comprises the U-dimensional spreading codes for each individual of V users), $\boldsymbol{A}$ is a $V \times V$ diagonal matrix which comprises the amplitude for each user. In fact, the amplitude shows the fading of the transmitted signal (the degree of fading represents how far the user is from the receiver). Finally, $\boldsymbol{n}_t$ stands for the known AWGN added to the U received chips $\boldsymbol{y}_t$ as time progresses.

At the receiver, the original signal $x_t(i)$ of the i^{th} user is needed to recover as follows:

$$\hat{x}_t(i) = \text{sgn}(\boldsymbol{w}_i^\top \boldsymbol{y}_t), \tag{2}$$

where $\boldsymbol{w}_i$ is the matched filter for the i^{th} user. Although the optimal solution of $\boldsymbol{w}_i$ is nonlinear, this vector can be estimated by using the MMSE estimation method as below:

$$\boldsymbol{w}_i^* = \arg\min_{\boldsymbol{w}_i} \mathbb{E}\left[(x_t(i) - \boldsymbol{w}_i^\top \boldsymbol{y}_t)^2\right] = \boldsymbol{C}_{yy}^{-1}\boldsymbol{C}_{yx}, \tag{3}$$

where $\boldsymbol{C}_{yy} = \mathbb{E}\left[\boldsymbol{y}_t \boldsymbol{y}_t^\top\right]$ is the auto-correlation of the received vectors, and $\boldsymbol{C}_{yx} = \mathbb{E}\left[\boldsymbol{y}_t x_t(i)\right]$ is the cross-correlation between the received vectors and original ones. Equation (3) is known to be the decentralized MMSE estimation and can be solved without the awareness of the spreading sequences of the other users. However, the problem of this solution is the huge size of the required training set and the high computational complexity regarding the matrix inverse.

D. Bui is with the Computer Engineering Department, Kyung Hee University, Suwon 446-701, Korea (e-mail: mao.bui@khu.ac.kr).
S. Lee is with the Computer Engineering Department, Kyung Hee University, Suwon 446-701, Korea (e-mail: sylee@oslab.khu.ac.kr).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LCOMM.2016.2620430, IEEE Communications Letters

2.2 MUD derivation of GPR

Let $\boldsymbol{\Phi} = [\boldsymbol{\phi}(\boldsymbol{y}_1), \boldsymbol{\phi}(\boldsymbol{y}_2)], \ldots, \boldsymbol{\phi}(\boldsymbol{y}_n)]$ denote the vector of nonlinear mapping into the higher dimensional space for the received signals, and $\boldsymbol{\phi}(\cdot)$ is the corresponding mapping function. Conditioning the original signal vector $\boldsymbol{x}(i)$ on the received signal vector $\boldsymbol{Y} = [\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_n]$, the distribution of $\boldsymbol{x}(i)$ is as follows:

$$p(\boldsymbol{x}(i)|\boldsymbol{Y}, \boldsymbol{w}) = \mathcal{N}(\boldsymbol{x}(i)|\boldsymbol{\Phi}^\top \boldsymbol{w}, \sigma^2_{noise}\, \boldsymbol{I}_n), \tag{4}$$

where σ_{noise} and $\sigma_{\boldsymbol{w}}$ are the standard deviations of the noise and the match filter, respectively; $\boldsymbol{I}_n$ is the identity matrix of size n. The aforementioned matched filter $\boldsymbol{w}$ is assumed to be a zero-mean Gaussian random variable. Hence, the distribution of $\boldsymbol{w}$ can be calculated as $p(\boldsymbol{w}) = \mathcal{N}(\boldsymbol{w}|0, \sigma^2_{\boldsymbol{w}}\, \boldsymbol{I}_n)$. By engaging Bayes' rule on (4), the posterior of $\boldsymbol{w}$ is calculated as follows:

$$p(\boldsymbol{w}|\boldsymbol{x}(i), \boldsymbol{Y}) = \frac{p(\boldsymbol{w})p(\boldsymbol{x}(i)|\boldsymbol{w}, \boldsymbol{Y})}{p(\boldsymbol{x}(i)|\boldsymbol{Y})}. \tag{5}$$

Theoretically, (3) can be transformed into a nonlinear form by using the maximum a posteriori (MAP) estimation of the random variable $\boldsymbol{w}$. The transformation is represented as below:

$$\boldsymbol{w}^* = \arg\min_{\boldsymbol{w}} \left\{||\boldsymbol{x}(i) - \boldsymbol{\Phi}^\top \boldsymbol{w}||^2 + \lambda||\boldsymbol{w}||^2\right\}, \tag{6}$$

where λ equals to $(\sigma_{noise}/\sigma_{\boldsymbol{w}})^2$. The term $\lambda||\boldsymbol{w}||^2$ is incorporated into the MAP as a regularizer to bypass the over-fitting issue. By finding $\boldsymbol{w}^*$, the estimation of the original signal $\hat{x}(i)$ can be obtained as follows:

$$\hat{x}(i) = \boldsymbol{k}^\top \mathbf{P}^{-1} \boldsymbol{x}(i), \tag{7}$$

where $\boldsymbol{k} = [k(\boldsymbol{y}, \boldsymbol{y}_1), k(\boldsymbol{y}, \boldsymbol{y}_2), \cdots, k(\boldsymbol{y}, \boldsymbol{y}_n)]$ with $k(\boldsymbol{y}_i, \boldsymbol{y}_j) = (\boldsymbol{\phi}(\boldsymbol{y}_i)^\top \boldsymbol{\phi}(\boldsymbol{y}_j))$ is the kernel function of the above nonlinear transformation; $\mathbf{P} = \mathbf{K} + \sigma^2_{noise}\, \mathbf{I}$ (where $\mathbf{K}$ is the covariance matrix with $\mathbf{K}_{ij} = k(\boldsymbol{y}_i, \boldsymbol{y}_j)$). After having the kernel function, (7) can be solved without inverting the matrix $\mathbf{P}$ by using the improved fast Gauss transform (IFGT) method [4]. It is worth noting that by applying the IFGT, the complexity of the solution for (7) drops to $O(n)$. The detail of IFGT implementation can be found in the original paper [4] and not be in the scope of this letter.

Due to the high priority of processing speed, the square exponential (SE) kernel function is engaged to compute the estimation of the original signals instead of the Matérn-class kernels [5] [6]. The SE kernel function is defined as below:

$$k(\boldsymbol{y}_i, \boldsymbol{y}_j) = \theta_1^2 \exp\left(-\frac{||\boldsymbol{y}_i - \boldsymbol{y}_j||^2}{2\theta_2^2}\right), \tag{8}$$

where θ_1 is an output-scale amplitude and θ_2 is a time-scale of $\boldsymbol{y}$ from one moment to the next. The set $\boldsymbol{\theta} = \{\theta_1, \theta_2\}$ is known as the set of hyper-parameters. The estimation of $\boldsymbol{\theta}$ can be calculated as below:

$$\boldsymbol{\theta}^* = \arg\max_{\boldsymbol{\theta}} p(\boldsymbol{x}(i)|\boldsymbol{Y}, \boldsymbol{\theta}). \tag{9}$$

Theoretically, the hyper-parameter set $\boldsymbol{\theta}^*$ can be estimated *via* minimizing the following negative marginal log likelihood $-\log p(\boldsymbol{x}(i)|\boldsymbol{Y}, \boldsymbol{\theta})$:

$$-\log p(\boldsymbol{x}(i)|\boldsymbol{Y}, \boldsymbol{\theta}) = \frac{1}{2}\boldsymbol{x}^\top(i)\mathbf{P}^{-1}\boldsymbol{x}(i) + \frac{1}{2}\log|\mathbf{P}| + \frac{n}{2}\log(2\pi). \tag{10}$$

In this equation, due to the high complexity of calculating the matrix inverse $\mathbf{P}^{-1}$, developing the approximation method is needed instead of finding exact values.

Analytically, in (10), the dominant computation focuses on two terms: the data-fit term [5] denoted by $\boldsymbol{x}^\top(i)\mathbf{P}^{-1}\boldsymbol{x}(i)$ and the log determinant $\log|\mathbf{P}|$. Solving these terms costs $O(n^3)$ for computational complexity [7]. Obviously, this drawback is a burden for DS-CDMA system. To solve the problem, we propose a complexity reduction method to significantly accelerate the calculation process. By equipped with the proposed method, the computational complexity can be dropped to $O(n \log n)$.

2.3 Complexity reduction

Our complexity reduction method is a combination of three techniques: the fast Fourier transform (FFT), the law of log determinant and the stochastic gradient descent (SGD). Instead of putting effort in minimizing the above negative marginal log likelihood, we think of approximately minimizing the upper bound of this term [8]. First, a simplification derivation is needed to apply on the aforementioned terms to compact the equation. In order to do that, the law of log determinant [9] is engaged to calculate the log determinant $\hat{P}$ of the empirical covariance matrix $\hat{\mathbf{P}}$, whereby (10) is simplified to

$$-\log p(\boldsymbol{x}(i)|\boldsymbol{Y}, \boldsymbol{\theta}) = \frac{1}{2}\boldsymbol{x}^\top(i)\mathbf{P}^{-1}\boldsymbol{x}(i) + \frac{1}{2}\hat{P} + \frac{n}{2}\log(2\pi), \tag{11}$$

where $\hat{P}$ is calculated based on the empirical covariance matrix $\hat{\mathbf{P}}$ and a constant τ as follows:

$$\begin{aligned} &\hat{P} = \log|\hat{\mathbf{P}}| - \tau, \\ &\text{with} \\ &\hat{\mathbf{P}} = \frac{1}{n-1}\sum_{k=1}^{n}[\boldsymbol{x}_k(i) - \bar{\boldsymbol{x}}(i)][\boldsymbol{x}_k(i) - \bar{\boldsymbol{x}}(i)]^\top, \\ &\tau = \gamma(\frac{n}{2}) - \log(\frac{n}{2}), \end{aligned} \tag{12}$$

where $\gamma(\cdot)$ is the Digamma function and $\bar{\boldsymbol{x}}(i)$ is the mean of the empirical data. After a number of re-calculations, the term $\hat{P}$ converges to a constant according to the central limit theorem [10]. This convergence leads to a consequence that minimizing the above negative marginal log likelihood might involve approximately minimizing the following reduced negative marginal log likelihood (rMLL):

$$-\log p(\boldsymbol{x}(i)|\boldsymbol{Y}, \boldsymbol{\theta}) = \frac{1}{2}\boldsymbol{x}^\top(i)\mathbf{P}^{-1}\boldsymbol{x}(i). \tag{13}$$

It is worth noting that (11) and (12) are used to derive (13). The full proofs and derivations can be found in the original paper [9]. For real implementation, the calculation of these equations is not required. Practically, solving the matrix inverse of $\mathbf{P}$ in (13) is still very computationally expensive. Therefore, there would be an alternative method to achieve the same goal. Since the covariance matrix $\mathbf{P}$ is positive-definite, it is possible to do the transformation by using FFT. This technique aims to bring the computation from the spatial-temporal domain into the frequency domain. It is worth noting that the cost of FFT is just $O(n \log n)$. Obviously, this cost is much better than the traditional method.

First, the SE kernel $k(\boldsymbol{y}_i, \boldsymbol{y}_j)$ in (8) needs to be re-written in Fourier transform representation [11] as shown below:

$$\mathcal{F}_{SE}(\omega) = \theta_2\theta_1^2\sqrt{2\pi}\ \exp\left(-2\pi^2\omega^2\theta_2^2\right), \tag{14}$$

where ω is the frequency representation of the received signal y in the periodic domain. Under this domain, (13) can be restated as follows:

$$-\log p(\boldsymbol{x}(i)|\boldsymbol{Y}, \boldsymbol{\theta}) = \frac{1}{2}\boldsymbol{x}_\circ^\top(i)\boldsymbol{\Psi}\boldsymbol{x}_\circ(i), \tag{15}$$

where $\boldsymbol{\Psi}$ is the function that generates $\mathbf{Q} = \mathbf{P}^{-1}$, $\boldsymbol{x}_\circ(i)$ denotes the data vector in the periodic domain. Subsequently, the Parseval theorem [12] [8] and the Fourier transform are applied to derive (15) as follows:

$$\mathcal{F}_{rMLL}(\theta) = \mathcal{F}\big(-\log p(\boldsymbol{x}(i)|\boldsymbol{y}, \boldsymbol{\theta})\big) = \frac{1}{2n}\tilde{\boldsymbol{x}}^\top(i)\widetilde{\boldsymbol{\Psi} * \boldsymbol{x}_\circ}(i), \tag{16}$$

where tilde sign and star sign denote a Fourier transform and a convolution, respectively. In the next step, by continuously applying the convolution theorem with regard to the convolution theorem's constraint: $\boldsymbol{\Psi}\mathcal{F}_{SE} \equiv 1$, the final form of the rMLL can be represented as follows:

$$\mathcal{F}_{rMLL}(\theta) = \frac{1}{2n}\sum_k \tilde{\Psi}_k * \tilde{x}_k^2(i) = \frac{1}{2n}\sum_k \frac{\tilde{x}_k^2(i)}{\mathcal{F}_{SE}(\omega_k)}, \tag{17}$$

where Ψ_k is the corresponding function of $\mathcal{F}_{SE}$ at the frequency ω_k with regard to the aforementioned constraint. By this form of (17), the set $\boldsymbol{\theta}$ of the hyper-parameters is estimated by using the gradient-based technique. In this case, the stochastic gradient descent (SGD) is chosen due to the properties of fast convergence and less sensitive to the local optima [13]. To integrate the SGD, the partial derivatives of (17) are required for each hyper-parameter. These equations are given by:

$$\begin{aligned}
\frac{\partial}{\partial\theta_2}\mathcal{F}_{rMLL} &= \tilde{x}_k^2(i)\exp\left(2\pi^2\theta_2^2\omega^2\right)\left(\frac{2\sqrt{2}\pi^{3/2}\omega^2}{\theta_1^2} - \frac{1}{\sqrt{2\pi}\theta_2^2\theta_1^2}\right)\\
\frac{\partial}{\partial\theta_1}\mathcal{F}_{rMLL} &= -\frac{\sqrt{\frac{2}{\pi}}\tilde{x}_k^2(i)\exp\left(2\pi^2\theta_2^2\omega^2\right)}{\theta_2\theta_1^3}.
\end{aligned} \tag{18}$$

Subsequently, an updating process is issued to update the hyper-parameters to the corresponding convergent points. This process is represented as below:

$$\begin{aligned}
\theta_2^{(k)} &\leftarrow \theta_2^{(k-1)} + \alpha(k)\frac{\partial}{\partial\theta_2^{(k-1)}}\mathcal{F}_{rMLL},\\
\theta_1^{(k)} &\leftarrow \theta_1^{(k-1)} + \alpha(k)\frac{\partial}{\partial\theta_1^{(k-1)}}\mathcal{F}_{rMLL},
\end{aligned} \tag{19}$$

where $\alpha(k) = {}^1\!/_{(k+1)}$ is the Robbins-Monroe decay function with regard to the k^{th} iteration. Clearly, the computation can be done without the matrix inverse. By the end of the proposed method, the required set $\boldsymbol{\theta}^*$ of the hyper-parameters is obtained within the computational complexity of $O(n\log n)$.

3 Performance Evaluation

In the experiments, we plan to evaluate some typical metrics in communications, namely the bit error rate (BER) and the signal to noise ratios (SNRs). The target system is of synchronous DS-CDMA type with 8 users spreading by the Gold sequences. Particularly, these binary sequences are generated with the length of 31. The powers of all users are equal with SNR = 4dB. The channel model is as follows:

$$H(z) = 0.4 + 0.9z^{-1} + 0.4z^{-2}. \tag{20}$$

For evaluation purpose, the traditional Gaussian process regression (traditional GP), the MMSE estimation and the proposed method are performed and compared together. Initially, the hyper-parameters θ_1 and θ_2 of GPR-family are set to 0. By the end of the learning process, the values of θ_1 and θ_2 are updated to 0.6782329 and 6.782329, respectively. Note that the MSE threshold is limited to $\epsilon = 0.1$. With this threshold, the hyper-parameters need around 4 to 5 iterations to reach the above values.

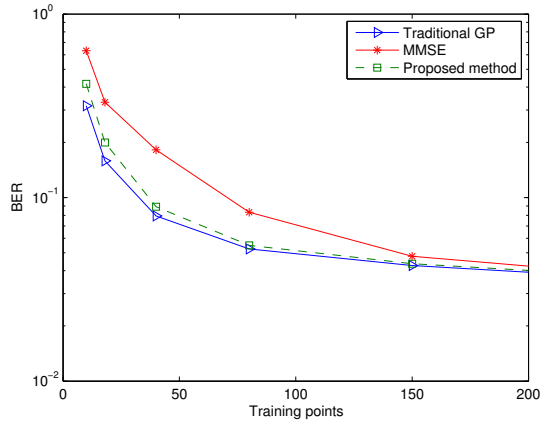
In Figure $1a$, a series of experiments are conducted in ascending order of size of the training dataset. For each experiment, the BER is computed for 10^6 bits. Obviously, the result of proposed method is very analogous to the traditional GP. It is worth mentioning that the GPR-family outperforms the MMSE estimation in terms of BER performance, especially when the size of training dataset is small. When the number of training points increases dramatically, the results of three approaches come close together. However, when training size skyrockets, the processing rate of the proposed method outperforms the remaining approaches as a consequence of lower complexity level. In order to make the conclusion more reasonable, the benchmark of completion time is conducted on a larger dataset with more than 3000 training points. This dataset is a subset of Google traces dataset [14]. The simulation is implemented in Python on a minimal CentOS system with no algorithmic change. Critically, the significant improvement of the proposed method can be seen in Figure $1b$.

In Figure $1c$ and $1d$, the relationship between the BER and the SNRs is depicted *via* the tests on 100 and 200 training points, respectively. Based on these tests, a conclusion can be made that the proposed method provides a very close result to the traditional GP. Only a small gap exists between the GPR approaches due to the error in the hyper-parameters approximation. Even in that case, the result of proposed method is still much better than the MMSE estimation.

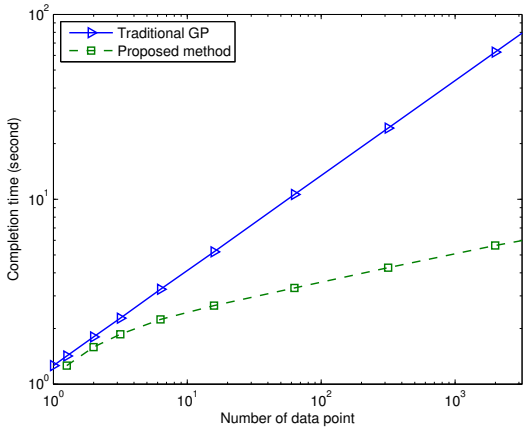
4 Conclusion

In this research, we propose an approach to reduce the computational complexity of the hyper-parameters estimation of GPR, which is mainly used to solve the issue of multiuser detection in DS-CDMA system. Previously, the contribution of GPR to the field is limited because of the high complexity. Therefore, by developing the complexity reduction method, we believe that the enhancement would innovate the development of Gaussian process-based applications to deal with the challenges in communication systems. For future works, the parallelism is also considered to be the next step to optimize our approach.

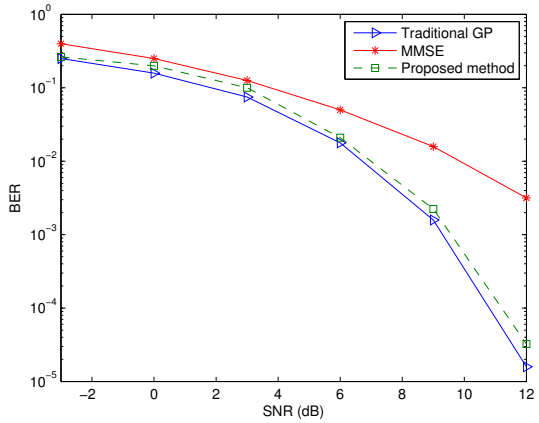
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LCOMM.2016.2620430, IEEE Communications Letters



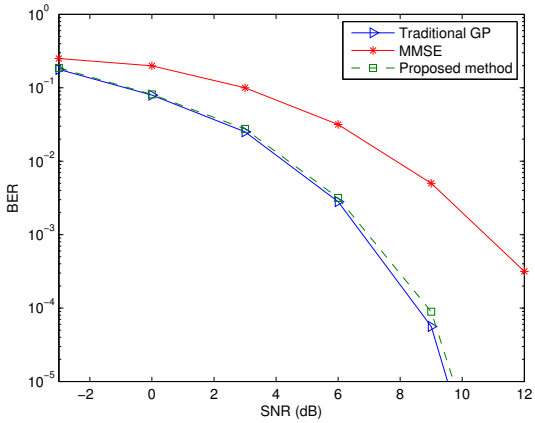
(a) Relationship between BER and the number of training points (SNR=4dB).



(b) Completion time benchmark (lower is better).



(c) BER vs SNRs on 100 training points.



(d) BER vs SNRs on 200 training points.

Fig. 1: Performance evaluation of proposed method

ACKNOWLEDGMENTS

This research was supported by the MSIP, Korea, under the G-ITRC (IITP-2015-R6812-15-0001) and ITRC ((IITP-2015-(H8501-15-1015) and (Grants No. C0395816) which was supported by Korea Small and Medium Business Administration in 2016. Corresponding author is Prof. Sungyoung Lee.

REFERENCES

[1] A. Duel-Hallen, J. Holtzman, and Z. Zvonar, "Multiuser detection for CDMA systems," *Personal Communications, IEEE*, vol. 2, no. 2, pp. 46–58, 1995.

[2] S. V. Halunga and N. Vizireanu, "Performance evaluation for conventional and mmse multiuser detection algorithms in imperfect reception conditions," *Digital Signal Processing*, vol. 20, no. 1, pp. 166–178, 2010.

[3] J. J. Murillo-Fuentes and F. Perez-Cruz, "Gaussian process regressors for multiuser detection in DS-CDMA systems," *IEEE Transactions on Communications*, vol. 57, no. 8, pp. 2339–2347, 2009.

[4] C. Yang, R. Duraiswami, and L. S. Davis, "Efficient kernel machines using the improved fast Gauss transform." in *NIPS*, pp. 1561–1568, 2004.

[5] C. Rasmussen and C. Williams, *Gaussian Processes for Machine Learning*, ser. Adaptive Computation And Machine Learning. MIT Press, 2005. [Online]. Available: http://www.gaussianprocess.org/gpml/chapters/

[6] A. Chiumento, M. Bennis, C. Desset, A. Bourdoux, L. Van der Perre, and S. Pollin, "Gaussian process regression for CSI and feedback estimation in LTE," in *Communication Workshop (ICCW), 2015 IEEE International Conference on*. IEEE, 2015, pp. 1440–1445.

[7] K. Chalupka, C. K. I. Williams, and I. Murray, "A framework for evaluating approximation methods for Gaussian process regression." *CoRR*, vol. abs/1205.6326, 2012.

[8] E. Rodner, A. Freytag, P. Bodesheim, and J. Denzler, "Large-scale Gaussian process classification with flexible adaptive histogram kernels." in *ECCV (4)*, ser. Lecture Notes in Computer Science, A. W. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, and C. Schmid, Eds., vol. 7575. Springer, 2012, pp. 85–98.

[9] T. T. Cai, T. Liang, and H. H. Zhou, "Law of log determinant of sample covariance matrix and optimal estimation of differential entropy for high-dimensional Gaussian distributions," *Journal of Multivariate Analysis*, vol. 137, pp. 161–172, 2015.

[10] R. Gallager, *Stochastic Processes: Theory for Applications*. Cambridge University Press, 2013. [Online]. Available: http://books.google.co.kr/books?id=CGFbAgAAQBAJ

[11] P. Sollich and C. K. I. Williams, "Understanding Gaussian process regression using the equivalent kernel." in *Deterministic and Statistical Methods in Machine Learning*, ser. Lecture Notes in Computer Science, J. Winkler, M. Niranjan, and N. D. Lawrence, Eds., vol. 3635. Springer, 2004, pp. 211–228. [Online]. Available: http://dblp.uni-trier.de/db/conf/dsmml/dsmml2004.html

[12] J. de Baar, R. Dwight, and H. Bijl, "Speeding up Kriging through fast estimation of the hyperparameters in the frequency-domain." *Computers & Geosciences*, vol. 54, no. 0, pp. 99–106, 2013.

[13] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004. [Online]. Available: http://books.google.co.kr/books?id=mYm0bLd3fcoC

[14] What is Google traces?, https://github.com/google/cluster-data/, accessed: 2016-01-21.