Data-Driven Geometric Programming for System-Level Performance Optimization

Soihem Gonmei
Department of Computer Science and Engineering
Seoul National University of Science and Technology
Seoul, South Korea
soihemgonmei@seoultech.ac.kr

Junhwan Lee
Electronics and Telecommunications Research Institute
Daejeon, South Korea
junhwanlee@etri.re.kr

Taesoo Kwon
Department of Computer Science and Engineering
Seoul National University of Science and Technology
Seoul, South Korea
tskwon@seoultech.ac.kr

***Abstract*—More than often, the output of a comprehensive network-wide performance modeling is a non-linear and non-convex function of the input data. To estimate the non-linear relationship of such a procedure, this paper employs a data-driven methodology. By formulating the non-linear objective as a geometric program, we leverage Levenberg-Marquardt algorithm to fit a convex log-sum-exponential function to data obtained through system-level simulation. Focusing on the downlink energy efficiency in mmWave cellular networks, we formulate a convex optimization problem and numerically obtain the optimal BS density and transmit power.**

***Index Terms*—Regression, Geometric Programming, Energy Efficiency, mmWave, Levenberg-Marquardt algorithm (LMA)**

I. Introduction

Ultra-dense networks (UDNs) can leverage mmWave and terahertz bands, deploying numerous small cells to meet surging mobile data demands. Yet, energy consumption remains a pivotal concern, driving the focus on enhancing energy efficiency (EE) in cellular networks [1]. The base station (BS) stands as a prime contributor to energy consumption. Optimizing BS density and transmit power emerges as a viable avenue for bolstering EE.

To curtail BS energy usage, strategies like switching off BS and micro BS deployment have emerged. Geometric programming (GP), noted for its adeptness in handling non-linear objectives and constraints, has found application in communication systems design [2], spanning power control [3], cost optimization [4], and EE enhancement [5]. Despite these advancements, much prior work relies on stochastic geometry, confined to simplified network settings. Studies such as [6] explore logistic function coverage approximations. Numerous works propose tools to fit data and create convex functional models. [7] introduces an efficient least-squares partition algorithm for data fitting with max-affine functions, paralleling [8]'s method for max-monomial functions. [9] proposes two convex function classes for convex regression. Fitting data with posynomial models is done in [10] and [11].

This paper investigates a data-driven method to model various network setups, capturing the performance metric's relationship with network parameters. We demonstrate effective convex approximation of EE data. Using two prevalent non-linear least squares algorithms, the Levenberg-Marquardt algorithm (LMA) and the Trust-Region-Reflective (TRR), we compare their fitting performance.

II. System Model

Consider a mmWave operated downlink multicell network where the BSs and user equipments (UEs) are spatially distributed according to a homogeneous Poisson Point Process, Φ_b and Φ_u with density λ_b and λ_u, respectively. All BSs transmit the same power P, commonly adjusted among them, and all BSs and UEs are equipped with a single antenna. Each UE is served by the BS with the strongest received signal. For a UE, the probability that a BS at distance r_i is line of sight (LoS) is $e^{-\beta r_i}$, where $1/\beta = 141.4\ m$, is the blockage density in urban area. If k-th BS is the serving BS, then the signal to interference plus noise ratio (SINR), γ is expressed as:

$$\gamma = \frac{P h_k G_b G_u ||r_k||^{-\alpha_{\mathrm{L/N}}}}{\sigma^2 + \sum_{i, r_i \in \phi_b \setminus r_k}^{K} P h_i \psi_i ||r_i||^{-\alpha_{\mathrm{L/N}}}}, \tag{1}$$

where $h_i \sim \exp(1)$ accounts for Rayleigh fading gain, σ^2 is the noise power, and $\psi_i \in \{G_b G_u, G_b g_u, g_b G_u, g_b g_u\}$, where G_b, g_b, G_u and g_u are the main and side lobe antenna gains for BS and UE, respectively. Path loss exponents for LoS and non-line of sight (NLoS) are α_{L} and α_{N}, when "L" and "N" denote LoS and NLoS, respectively. For a given SINR threshold θ, the areal throughput and EE are defined as:

$$\mathrm{T}(\lambda_b, P, \theta) = \lambda_b \log_2(1+\theta) \Pr[\gamma \geq \theta], \tag{2}$$

$$EE = \frac{\mathrm{T}(\lambda_b, P, \theta)}{\lambda_b \mathrm{P}_{\mathrm{tot}}(P)} = \frac{\mathrm{T}(\lambda_b, P, \theta)}{\lambda_b \rho \left(P_C + \frac{P}{\eta}\right)}, \tag{3}$$

where $\mathrm{P}_{\mathrm{tot}}(P)$ = total power consumption of BS, $P_C = 2.1\mathrm{W}$ is the circuit power per antenna, $\eta = 0.08$ is the power amplifier efficiency of BSs and $\rho = 1.21$ is the power loss rate of the DC-DC converter, current supply and cooling [11].

III. Overview of geometric programming

To model the network with GP [4], we parameterize a GP compatible function through convex regression. In GP, the objective and constraints can only be composed of monomial and posynomial functions. A posynomial is a function $g(\mathbf{x}): \mathbb{R}^n_{++} \rightarrow \mathbb{R}_{++}$ of the form $g(\mathbf{x}) = \sum_{k=1}^{K} c_k \prod_{j=1}^{n} x_j^{a_{jk}}$ where $c_k \in \mathbb{R}_{++}$ and $\mathbf{a}_k = (a_{1k}, ..., a_{nk}) \in \mathbb{R}^n$. When $K = 1$, posynomial (g_i) becomes monomial (h_i). GP is formulated as:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & g_0(\mathbf{x}) \\ \text{subject to} \quad & g_i(\mathbf{x}) \leq 1, \ i = 1, ..., m \\ & h_i(\mathbf{x}) = 1, \ i = 1, ..., n \end{aligned} \tag{4}$$

The transformation $u_i = \log x_i$ and $b_k = \log(c_k)$ converts GP into convex optimization problem and the posynomial function $g(\mathbf{x})$ becomes log-sum-exp function $f_0(\mathbf{u}) = \log\left(\sum_{k=1}^{K_0} e^{a_{0k}^T \mathbf{u} + b_{0k}}\right)$.

IV. EE Performance and Data Fitting

Given a set of data points which is multivariate, $\mathcal{D} = \{(\mathbf{x}_i, w_i) | i = 1, ..., m\}$, we are interested in modelling these data points with posynomial function which is GP-compatible. Our fitting procedure involves logarithmic transformation of the multivariate data as $(\mathbf{u}_i, w_i) = (\log \mathbf{x}_i, \log y_i)$, and fitting a convex function to the resulting data, where $\log \mathbf{x}_i = (\log x_1, ..., \log x_n)$ is the vector containing the log of the decision variables for the ith data point. Therefore, the original data must be strictly positive because of the log transformation. Based on (3), the EE optimization for the multicell network is formulated as follows,

$$\begin{aligned} \underset{\lambda_b, P}{\text{maximize}} \quad & \text{EE}(\lambda_b, P, \theta) \\ \text{subject to} \quad & \lambda_{min} \leq \lambda_b \leq \lambda_{max} \\ & P_{min} \leq P \leq P_{max} \end{aligned} \tag{5}$$

This aims to solve for the optimal λ_b and P that maximize EE under the given constraints. The data obtained from simulation $\{\lambda_b, P, EE\}$ is log transformed, denoted as $\{\tilde{\lambda}_b, \tilde{P}, \widetilde{EE}\}$. We consider that SINR threshold θ is fixed meaning that all users have the same rate $\log_2(1+\theta)$ bps/Hz if transmission is successful, i.e., $\gamma \geq \theta$. Now we can fit the log-transformed data with the log-sum-exp function.

A. Fitting model parameters

This section describes the fitting process. We use the log-sum-exponential function as our convex fitting function. Given m data points $(\mathbf{u}_i, w_i) \in \mathbb{R}^n \times \mathbb{R}$, we aim to minimize the least squared error: $F = \sum_{i=1}^{m} (f_0(\mathbf{u}_i; \beta) - w_i)^2$ where f_0 is our fitting function and $\beta \in \mathbb{R}^n$ is a vector containing the function parameters $\mathbf{a}$ and $\mathbf{b}$. The least squares problem is non-convex, i.e., it can have multiple local minima.

B. Levenberg-Marquardt algorithm

The LMA, Algorithm 1 [12], uses a search direction that is a solution of the linear set of equations: $(\mathbf{J}^T\mathbf{J} + \lambda diag(\mathbf{J}^T\mathbf{J}))\delta = -\mathbf{J}^T\mathbf{r}$, where $\mathbf{J}$ is the Jacobian of the objective function, F and $\mathbf{r}(\beta) = f_0(X; \beta) - Y$ is the residual vector. D adjusts the trust region, Δ for next iteration after parameter update, δ. The regularization term, λ smoothly transitions between the steepest descent method and Gauss-Newton method. When λ is small, LMA behaves more like the Gauss-Newton method, leading to faster convergence in well-conditioned regions. And when λ is large, LMA behaves more like the steepest descent method, providing better convergence in ill-conditioned regions. Hence, the LMA combines the advantages of the two algorihms. The second algorithm, Trust-Region-Reflective (TRR) operates within a trust region. Calculation of the step between iterates requires the solutions of a problem of the form: min $\Psi(\omega) : ||\omega|| \leq \Delta$, where Ψ represents a local model to the objective function, and Δ is the maximum allowable step length in the parameter space.

TABLE I

System Parameters and their values

Parameter	Value
BS transmit power P	5 dBm to 55 dBm
BS density λ_b	1 km^{-2} to 1000 km^{-2}
UE density λ_u	$(100/\pi)$ km^{-2}
Pathloss exponent, α_{L}, α_{N}	$2, 4$
Carrier frequency f_c	28 GHz
BS antenna gains, G_b, g_b	18 dB, -2 dB
UE antenna gains, G_u, g_u	0 dB, 0 dB
Frequency Bandwidth, W	100 MHz
Noise Power, σ^2	-174 dBm/Hz + $10\log_{10} W$ + 10 dB

Algorithm 1 Levenberg-Marquardt iteration

Given $\Delta_k > 0$, find $\lambda_k \geq 0$ such that
if $(\mathbf{J}^T\mathbf{J} + \lambda diag(\mathbf{J}^T\mathbf{J}))\delta = -\mathbf{J}^T\mathbf{r}$ **then**
  either $\lambda_k = 0$ and $||D_k\delta_k|| \leq D_k$
  or $\lambda_k > 0$ and $||D_k\delta_k|| = \Delta_k$
end if
if $||F(x_k + \delta_k)|| < ||F(x_k)||$ **then**
  set $x_{k+1} = x_k + \delta_k$ and evaluate $\mathbf{J}_{k+1}$
else set $x_{k+1} = x_k$ and $\mathbf{J}_{k+1} = \mathbf{J}_k$
end if
choose Δ_k and D_{k+1}

V. Numerical Results and Conclusions

The values of system parameters for the Monte Carlo simulations of the network are listed in Table I.

The nonlinear least squares problem is randomly initialized multiple times. For fair comparison, the same initial parameters in the two algorithms are used. The function tolerance is set to 10^{-6}. Table II summarizes a typical performance. "Iteration" is the number of repetition until the change in the sum of squares w.r.t, initial value is less than function tolerance. "Function-count" is the total times the objective function is assessed, informing about computational efficiency. "Residual" is the variance between observed and predicted values in the model.

Fig. 1 and 2 illustrate the fitted curves on the simulation data (square dots) using LMA and TRR. Both algorithms demonstrate effective fitting. In our network context, LMA exhibits quicker convergence on average and involves fewer function evaluations. It is worth noting that while both may not

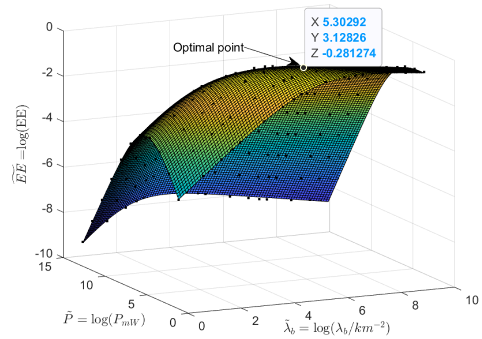

Fig. 1. Fitted log-sum-exp function on the log-transformed data $(\tilde{\lambda}_b, \tilde{P}, \widetilde{EE})$ with $K = 7$ affine terms and residual = 0.502205 after 42 LMA iterations.

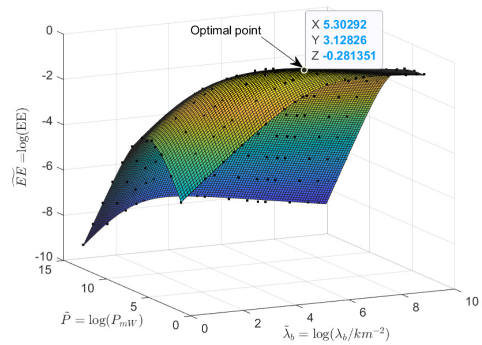

Fig. 2. Fitted log-sum-exp function on the log-transformed data $(\tilde{\lambda}_b, \tilde{P}, \widetilde{EE})$ with $K = 7$ affine terms and residual = 0.502557 after 95 TRR iterations.

necessarily converge to the same optimum, the final residuals are similar, resulting in near-identical graph shapes and nearly equal optimal solutions. The fitted functions in Fig. 1 and Fig. 2 are parameterized as follows:

$$\begin{aligned} f_{0\ LMA}(\tilde{\lambda}_b, \tilde{P}) = &-\log(e^{-3.4677-1.1814\tilde{\lambda}_b+0.9754\tilde{P}}+ \\ & e^{6.1248-0.9965\tilde{\lambda}_b-0.9019\tilde{P}} + e^{0.4250-2.4019\tilde{\lambda}_b-2.3266\tilde{P}}+ \\ & e^{2.1369-0.4621\tilde{\lambda}_b-1.2477\tilde{P}} + e^{-0.7696+0.1492\tilde{\lambda}_b-0.0018\tilde{P}}+ \\ & e^{-6.2438+0.1935\tilde{\lambda}_b+1.0004\tilde{P}} + e^{2.3149-1.0957\tilde{\lambda}_b+0.0066\tilde{P}}) \end{aligned} \tag{6}$$

$$\begin{aligned} f_{0\ TRR}(\tilde{\lambda}_b, \tilde{P}) = &-\log(e^{-0.7881+0.1500\tilde{\lambda}_b+0.0008\tilde{P}}+ \\ & e^{2.3640-1.0979\tilde{\lambda}_b-0.0005\tilde{P}} + e^{6.1353-0.9903\tilde{\lambda}_b-0.9091\tilde{P}}+ \\ & e^{-6.2448+0.1934\tilde{\lambda}_b+1.0005\tilde{P}} + e^{0.7722-0.2986\tilde{\lambda}_b-1.0073\tilde{P}}+ \\ & e^{-3.4632-1.1812\tilde{\lambda}_b+0.9750\tilde{P}} + e^{25.0087-0.7991\tilde{\lambda}_b-22.0761\tilde{P}}) \end{aligned} \tag{7}$$

The final EE optimization problem is formulated below:

$$\begin{aligned} \max_{\tilde{\lambda}_b, \tilde{P}} \quad & -\log(\sum_{k}^{K} \exp\left(b_k + a_{k\tilde{\lambda}_b}\tilde{\lambda}_b + a_{k\tilde{P}}\tilde{P}\right)) \\ \text{subject to} \quad & \tilde{\lambda}_{min} \le \tilde{\lambda}_b \le \tilde{\lambda}_{max} \\ & \tilde{P}_{min} \le \tilde{P} \le \tilde{P}_{max} \end{aligned} \tag{8}$$

TABLE II
TYPICAL PERFORMANCES OF LMA AND TRR

Algorithm	Iteration	Function-count	Residual
LMA	59	1357	0.505634
	68	1555	0.502218
	42	970	0.502205
TRR	95	2112	0.518031
	79	1760	0.502225
	95	2112	0.502557

Through LMA and TRR fitting, we demonstrated the efficacy of data-driven modeling for GP. LMA showcased swifter convergence and fewer evaluations, though both methods yielded near-identical solutions. Future work will address the solution to the formulated GP.

ACKNOWLEDGMENT

This work was supported by Institute for Information & communications Technology Promotion(IITP) grant funded by the Korea government(MSIT) (No.2021-0-02102 , "Global Standard Collaboration on Vision and Wireless Channel Characteristics of 6G Wireless Mobile Communications")

REFERENCES

[1] X. Ge, B. Yang, J. Ye, G. Mao, C.-X. Wang, and T. Han, "Spatial spectrum and energy efficiency of random cellular networks," *IEEE Trans. Commun.*, vol. 63, no. 3, pp. 1019–1030, Mar. 2015.

[2] M. Chiang, "Geometric Programming for Communication Systems", *Foundations and Trends in Communications and Information Theory*, vol. 2, no. 1-2, pp. 1-154, 2005.

[3] M. Chiang, C. W. Tan, D. P. Palomar, D. O'neill, and D. Julian, "Power Control By Geometric Programming," *IEEE Trans. on Wireless Commun.*, vol. 6, no. 7, pp. 2640-2651, Jul. 2007.

[4] T. Kwon, S. W. Choi, and Y. -H. Shin, "A Comprehensive Design Framework for Network-Wide Cost Reduction in Random Access-Based Wireless IoT Networks," *IEEE Commun. Lett.*, vol. 23, no. 9, pp. 1576-1580, Sept. 2019.

[5] S. Cao, and F. Hou, "On the Mathematical Modeling and Optimization for the Energy Efficiency Performance of CSMA-NOMA Random Access Networks With Channel Inversion," *IEEE Trans. on Wireless Commun.*, vol. 22, no. 4, pp. 2867-2884, Apr. 2023.

[6] H. E. Hammouti, M. Ghogho, and S. A. Raza Zaidi, "A Machine Learning Approach to Predicting Coverage in Random Wireless Networks," *IEEE Globecom Workshops (GC Wkshps)*, Abu Dhabi, United Arab Emirates, 2018.

[7] A. Magnani and S. P. Boyd, "Convex piecewise-linear fitting," *Optim. Eng.*, vol. 10, no. 3, pp. 1–17, 2009.

[8] J. Kim, L. Vandenberghe, and C.-K. K. Yang, "Convex piecewise-linear modeling method for circuit optimization via geometric programming," *IEEE Trans. Comput. Aided Des. Integr. Circuits Syst.*, vol. 29, no. 11, pp. 1823–1827, Nov. 2010

[9] W. Hoburg, P. Kirschen, and P. Abbeel, "Data fitting with geometric-programming-compatible softmax functions," *Optim. Eng.*, vol. 17, pp. 897–918, 2016.

[10] W. Daems, G. Gielen, and W. Sansen, "Simulation-based generation of posynomial performance models for the sizing of analog integrated circuits," *IEEE Trans. Comput.-Aided Design Integr.*, vol. 22, no. 5, pp. 517–534, May 2003.

[11] C. Chen, J. Zhang, X. Chu, and J. Zhang, "On the Deployment of Small Cells in 3D HetNets With Multi-Antenna Base Stations," *IEEE Trans. on Wireless Commun.*, vol. 21, no. 11, pp. 9761-9774, Nov. 2022

[12] J. J. Moré, "The Levenberg-Marquardt algorithm: Implementation and theory," *Numerical Analysis*. Berlin, Germany: Springer, pp. 105–116, 1978.