Utility-Based Precoding Optimization Framework for Large Intelligent Surfaces

Emil Björnson*, Luca Sanguinetti†

*Department of Electrical Engineering (ISY), Linköping University, Linköping, Sweden (emil.bjornson@liu.se)
†Dipartimento di Ingegneria dell’Informazione, University of Pisa, 56122 Pisa, Italy (luca.sanguinetti@unipi.it)

Abstract—The spectral efficiency of wireless networks can be made nearly infinitely large by deploying many antennas, but the deployment of very many antennas requires new topologies beyond the compact and discrete antenna arrays used by conventional base stations. In this paper, we consider the large intelligent surface scenario where small antennas are deployed on a large and dense two-dimensional grid. Building on the heritage of MIMO, we first analyze the beamwidth and sidelobes on a large and dense two-dimensional grid. Building on the methodology in [4, Sec. 7.3.1], if we order the antennas row by row, starting with the antenna in the origin, the array response vector can be expressed as

\[ \mathbf{a}(\varphi, \theta) = \left[ e^{j\alpha_{1,1}} \ldots e^{j\alpha_{i,1}} \ldots e^{j\alpha_{N,M}} \right]^T \]  

where \( j \) is the imaginary unit and the phase-shift of the antenna at row \( n \) and column \( m \) is

\[ \alpha_{n,m} = \frac{2\pi d}{\lambda} \left( (m-1) \cos(\theta) \sin(\varphi) + (n-1) \sin(\theta) \right). \]

Note that \( \|\mathbf{a}(\varphi, \theta)\|^2 = NM \). When transmitting from the array, \( \mathbf{a}(\varphi, \theta) \) is the channel vector to a user located in the far-field in azimuth angle \( \varphi \) and elevation angle \( \theta \). The propagation distance should be larger than the Fraunhofer distance \( 2d^2 \max(M^2,N^2) \) to use this far-field model.

We consider two single-antenna users, which are located in the far-field of the array in the directions \( (\varphi_1, \theta_1) \) and \( (\varphi_2, \theta_2) \). We assume the angles are non-identical: either \( \varphi_1 \neq \varphi_2 \) and/or \( \theta_1 \neq \theta_2 \). The transmissions are precoded using the vectors \( \sqrt{\rho_i} \mathbf{v}_i \) for user \( i = 1,2 \). More precisely, \( \mathbf{v}_i \in \mathbb{C}^{NM} \) is the unit-norm precoding vector assigned to user \( i \) and \( \rho_i \) is the normalized transmit power, which represents the signal-to-noise ratio (SNR). Moreover, let \( s_1, s_2 \) denote the independent information-bearing signals with \( \mathbb{E}\{|s_1|^2\} = \mathbb{E}\{|s_2|^2\} = 1 \).

Fig. 1. A planar array with \( N \) rows and \( M \) antennas per row. If a plane wave is impinging from elevation angle \( \theta \) and azimuth angle \( \varphi \), the array response vector is \( \mathbf{a}(\varphi, \theta) \).
The received signal at user 1 is modeled as
\[ y_1 = \mathbf{a}^H(\varphi_1, \theta_1)(\sqrt{r_1}v_1s_1 + \sqrt{r_2}v_2s_2) + n_1 \] (3)
where \( n_1 \sim \mathcal{N}(0, 1) \) is the normalized receiver noise. The received signal at user 2 is achieved by switching the user indices and is therefore omitted. By treating interference as noise, the achievable spectral efficiency (SE) is \( \log_2(1 + \text{SINR}_1) \), where the signal-to-interference-and-noise ratio (SINR) is
\[ \text{SINR}_1 = \frac{\rho_1|\mathbf{a}^H(\varphi_1, \theta_1)v_1|^2}{\rho_2|\mathbf{a}^H(\varphi_1, \theta_1)v_2|^2 + 1}. \] (4)
To further analyze this expression, we will now consider two different precoding schemes: MR and ZF precoding.

### A. Maximum Ratio Transmission

The first precoding scheme is MR
\[ v_i = \frac{\mathbf{a}(\varphi_i, \theta_i)}{\|\mathbf{a}(\varphi_i, \theta_i)\|} \quad i = 1, 2 \] (5)
which is optimal in noise-limited scenarios \( \text{[8]} \). When substituting these precoding vectors into (4), we obtain
\[ \text{SINR}^{\text{MR}}_1 = \frac{\text{SNR}_1}{\text{SNR}_2 \cdot I_{12} + 1}, \] (6)
where
\[ I_{12} = \left| \frac{1}{NM} \sum_{n=1}^{N} \sum_{m=1}^{M} e^{j(\alpha_{nm} - \alpha_{n,m} - \theta_1)} \right| \] (7)
accounts for the interference generated by user 2 and \( \text{SNR}_i = NM \frac{\rho_i}{2} \) for \( i = 1, 2 \) represents the received SNR of user \( i \) in the absence of any interference. Observe that
\[ I_{12} = \left| \frac{1}{NM} \sum_{n=1}^{N} \sum_{m=1}^{M} e^{j2\pi\lambda/(n-1)\Omega} \right| \left| \frac{1}{M} \sum_{m=1}^{M} e^{j2\pi\lambda/(m-1)\Psi} \right| \] (8)
where
\[ \Omega = \sin(\theta_2) - \sin(\theta_1), \] (9)
\[ \Psi = \cos(\theta_2) \sin(\varphi_2) - \cos(\theta_1) \sin(\varphi_1). \] (10)

**Lemma 1.** For any integer \( N \geq 1 \) and real-valued \( A \),
\[ \sum_{n=1}^{N} e^{j2\pi(n-1)A} = \left\{ \begin{array}{ll} \sin(\pi A) & A \neq 0, \\ N, & A = 0. \end{array} \right. \] (11)
By using the above lemma, we can rewrite (8) as
\[ I_{12} = \left| \frac{1}{NM} \sin(\pi N \Omega / \lambda) \sin(\pi M \Psi / \lambda) \right| \] (12)

### B. ZF Precoding

ZF is the optimal precoding scheme in interference-limited scenarios \( \text{[8]} \) and there are two equivalent definitions: using a pseudo-inverse or an orthogonal projection matrix \( \text{[9]} \), Sec. 3.4.2. We consider the latter formulation for which
\[ v_i = \frac{\mathbf{w}_i}{\|\mathbf{w}_i\|} \quad \text{for } i = 1, 2 \] (13)
\[ w_1 = \left( \mathbf{I}_{NM} - \frac{1}{NM} \mathbf{a}(\varphi_2, \theta_2)\mathbf{a}^H(\varphi_2, \theta_2) \right) \mathbf{a}(\varphi_1, \theta_1) \] (14)
\[ w_2 = \left( \mathbf{I}_{NM} - \frac{1}{NM} \mathbf{a}(\varphi_1, \theta_1)\mathbf{a}^H(\varphi_1, \theta_1) \right) \mathbf{a}(\varphi_2, \theta_2). \] (15)
These vectors satisfy the ZF properties \( \mathbf{a}^H(\varphi_2, \theta_2)v_1 = 0 \) and \( \mathbf{a}^H(\varphi_1, \theta_1)v_2 = 0 \). Hence, substituting \( v_i \) into (4) yields
\[ \text{SINR}^{\text{ZF}}_1 = \frac{\rho_1|\mathbf{a}^H(\varphi_1, \theta_1)w_1|^2}{\|w_1\|^2} = \text{SNR}_1(1 - I^2_{12}). \] (16)
Interestingly, (16) contains the same components as (4) (\( \text{SNR}_1, I^2_{12}, \) and 1), but has a different structure. In (16), \( I^2_{12} \) should be interpreted as the performance loss due to the cancellation of the interference generated by user 2.

### III. System Analysis for Dense Arrays

We will now analyze the system above in the limit of infinitesimal antennas for a given array dimension, which represents an ideal LIS. More precisely, we fix the array’s horizontal length to \( L = Md \) and the vertical height to \( H = Nd \), and then we will let \( d \to 0 \). Each antenna has a physical size of \( d \times d \) but the effective size shrinks to \( d \cos(\varphi_i) \times d \cos(\theta_i) \) when observing it from the direction of user \( i \). The SNR per antenna reduces with the effective antenna area \( \mathcal{A} \), which in free-space propagation is modeled as
\[ \frac{\text{SNR}_i}{N} = \frac{q_i}{\sigma^2} \frac{d^2 \cos(\varphi_i) \cos(\theta_i)}{4 \pi \mathcal{A}} \] (17)
for \( i = 1, 2 \), where \( q_i \) is the unnormalized transmit power, \( \sigma^2 \) is the noise power, \( r_i \) is the distance to user \( i \). As indicated in (17), we denote the part that does not depend on \( d \) as \( p_i \).

#### A. Limiting SINRs

By using (17), we have that
\[ \text{SNR}_i = NM \frac{p_i}{\sigma^2} = p_i d^2 NM = p_i LH \quad \text{for } i = 1, 2, \] (18)
depends on the array area \( LH \) but not on the area \( d^2 \) of each antenna. However, the interference gain in (12) depends on \( d^2 \):
\[ I^2_{12} = 1 \frac{\sin(\pi H \Omega / \lambda) \sin(\pi L \Psi / \lambda)}{N M^2 \sin^2(\pi d \Omega / \lambda) \sin^2(\pi d \Psi / \lambda)}. \] (19)
By letting \( d \to 0 \) and utilizing that \( \sin(x) \approx x \) is a tight approximation as \( x \to 0 \), the following limit is obtained.

**Lemma 2.** If \( d \to 0 \), then
\[ I^2_{12,d=0} = \lim_{d \to 0} I^2_{12} = \text{sinc}^2 \left( \frac{H \Omega}{\lambda} \right) \text{sinc}^2 \left( \frac{L \Psi}{\lambda} \right) \] (20)
where \( \text{sinc}(x) = \sin(\pi x) / (\pi x) \) is the sinc-function.
The above limit is in general non-zero, which was expected since the spatial resolution of an array is known to depend on the aperture (i.e., length $L$ and height $H$) and not the antenna spacing; see for example, [3] Sec. 7.2.4, [4] Sec. 7.4.2, [5]. The two squared sinc-functions determine how large the interference is. Since the two users are assumed to have non-identical angles, we have $\Omega \neq 0$ and/or $\Psi \neq 0$, which implies that at least one of the sinc-functions can be small if the array is physically large. By using Lemma 2 the limiting SINRs with MR and ZF easily follow.

**Lemma 3.** The limiting SINRs with MR and ZF are:

\[
\text{SINR}_{1,d=0}^{\text{MR}} = \lim_{d \to 0} \text{SINR}_{1,d=0}^{\text{MR}} = p_1 L H \\
= p_2 L H \sin^2 \left( \frac{H \Omega}{\lambda} \right) \sin^2 \left( \frac{L \Psi}{\lambda} \right) + 1
\]

(21)

\[
\text{SINR}_{1,d=0}^{\text{ZF}} = \lim_{d \to 0} \text{SINR}_{1,d=0}^{\text{ZF}} = p_1 L H \left( 1 - \sin^2 \left( \frac{H \Omega}{\lambda} \right) \sin^2 \left( \frac{L \Psi}{\lambda} \right) \right).
\]

(22)

**B. Interference Gain: Beamwidth and Sidelobes**

We now analyze the interference gain when user 1 has $\varphi_1 = \theta_1 = 0$, while the interfering user 2 has $\theta_2 = 0$ but a varying azimuth angle $\varphi_2 \in [-\pi/2, \pi/2]$. From (9) and (10), we thus obtain $\Omega = 0$ and $\Psi = \sin(\varphi_2)$ such that (19) reduces to

\[
I_{12}^2 = \frac{d^2 \sin^2(\pi L \Psi/\lambda)}{L^2 \sin^2(\pi d \Psi/\lambda)}
\]

(23)

which shows that, for any given $\Psi$ and $L/\lambda$, the interference gain depends on $d$, where a smaller $d$ leads to smaller values. From Lemma 2 the limit is given by

\[
I_{12,d=0}^2 = \sin^2 \left( \frac{L \Psi}{\lambda} \right).
\]

(24)

The maximum of both (23) and (24) is achieved for $\varphi_2 = 0$, which makes $\Psi = 0$. To find the nulls of the interference gain, (23) and (24) are set equal to zero, which leads to

\[
\varphi_2 = \varphi_{2,n} = \pm \arcsin \left( \frac{L \Psi}{\lambda} \right) \approx \pm \frac{\lambda}{L} n = 1, 2, \ldots
\]

(25)

where the approximation holds for $L \gg \lambda n$. If we define the beamwidth as the angular distance between the first two nulls, it is approximately $2\lambda/L$. In line with classical results on the resolution of arrays [3] Sec. 7.2.4, the beamwidth is inversely proportional to the array length $L$, but independent of $d$.

The maxima of the sidelobes occur when the numerator of (23) and (24) attains its maximum; that is, $\sin(\pi L \Psi/\lambda) = \pm 1$. For $L \gg \lambda n$, it is approximately given by

\[
\varphi_2 = \varphi_{2,n}^\text{max} \approx \pm \frac{2n + 1}{2} \frac{\lambda}{L} n = 1, 2, \ldots
\]

(26)

By using the above approximation in (25) and (24), we obtain

\[
I_{12,d=0}^2 \varphi_2 = \varphi_{2,n}^\text{max} \approx \frac{d^2 \sin^2(\pi L \varphi_{2,n}^\text{max}/\lambda)}{L^2 \sin^2(\pi d \varphi_{2,n}^\text{max}/\lambda)} = \frac{1}{L^2 \sin^2 \left( \frac{2n+1}{2} \pi d / L \right)}
\]

(27)

From which it follows that the maximum of the first sidelobe (i.e., $n = 1$) of an ideal LIS is $(\frac{\lambda}{2\pi})^2 = -13.46$ dB weaker than the main lobe, irrespective of the surface length $L$. The difference between (27) and (28) depends on $d/\lambda$ and $n$, where smaller $d/\lambda$ and/or $n$ lead to smaller differences. As a rule-of-thumb, we can approximate (27) with (28) whenever $\pi d \varphi_{2,n}^\text{max}/\lambda \leq \pi^2/8$ since then $\sin(x) \approx x$ with an error below 10%. For $\varphi_{2,n}^\text{max} \in [-\pi/4, \pi/4]$, we obtain $d \leq \lambda/2$.

Fig. 2 shows the interference gain from a user located at $\theta_2 = 0$ and $\varphi_2 \in [-\pi/2, \pi/2]$, while the desired user is at $\varphi_1 = \theta_1 = 0$.

![Fig. 2](image_url)

Fig. 2. The interference gain from a user located at $\theta_2 = 0$ and $\varphi_2 \in [-\pi/2, \pi/2]$, while the desired user is at $\varphi_1 = \theta_1 = 0$.

**C. MR versus ZF Preoding**

We will continue evaluating the SE that is achieved by user 1 with MR and ZF. We use the asymptotic expressions in (21) and (22) with $L = H = 50\lambda$. User 1 is located at the angles $\varphi_1 = \theta_1 = 0$, while we vary both angles for user 2. We assume
D. Asymptotic Analysis with $L \to \infty$

Although ZF outperforms MR for a finite surface, the situation might change as the surface grows. We can let the surface grow large, for example, by letting $L \to \infty$. This limit is not practically achievable using our far-field channel model, but it is still accurate for very large arrays [11].
which depends on $\text{SNR}_k = \rho_k MN$ (i.e., the received SNR of user $k$ without interference) and a constant $b_k \geq 0$ given by

$$b_k = 1 - \frac{1}{NM} a^H(\varphi_k, \theta_k) A_k(\varphi_k, \theta_k)^{-1} A_k^H a(\varphi_k, \theta_k). \quad (34)$$

It remains to jointly optimize the SINRs by selecting the transmit powers under a total transmit power $Q$. From (37), we have that

$$\sum_{i=1}^{K} q_i = \sum_{i=1}^{K} \rho_i \frac{4\pi r_i^2 \sigma^2}{d^2 \cos(\theta_i) \cos(\varphi_i)} \leq Q. \quad (35)$$

To determine what is a good power allocation, we define an increasing utility function $U(x)$ and consider the following utility maximization problem:

$$\max_{\rho_1, \ldots, \rho_K} \sum_{i=1}^{K} U(\rho_i b_i) \quad (36)$$

subject to

$$\sum_{i=1}^{K} \rho_i c_i \leq Q. \quad (37)$$

Lemma 5. If $U(x)$ is differentiable and $U'(x) = \frac{d^2}{dx^2} U(x)$ is invertible, then the solution to (36) is

$$\rho_i = \frac{1}{b_i} \left[ U'\left( \frac{c_i}{\rho_i b_i} \right) \right]_+ \quad (38)$$

where $[.]_+$ replaces negative values with zero, and the parameter $\nu \geq 0$ is selected to achieve equality in (37).

Proof: Follows from adapting [2] Th. 3.16.

If $U(x) = \log(x)$, then we are maximizing the product of the SINRs, which is called proportional fairness [4]. We then have $U'(x) = 1/x$ and $U'^{-1}(y) = 1/y$, so that (38) becomes

$$\rho_i = \frac{1}{b_i} \left[ \frac{c_i}{\nu \log_c(2) c_i} \right]_+ = \frac{\nu}{c_i} = \frac{Q}{c_i K} \quad (39)$$

since $\nu = Q/K$ gives equality in the power constraint. This leads to an equal power allocation since $p_i = c_i c_i$. If $U(x) = \log_2(1 + x)$, then we are instead maximizing the sum SE. It follows that $U'^{-1}(y) = \frac{1}{\nu \log_2(2)} - 1$ and (38) becomes identical to conventional waterfilling [3]:

$$\rho_i = \frac{\nu}{\log_2(2) c_i} - \frac{1}{b_i} \quad (40)$$

If we want to maximize the harmonic mean of the SINRs, $K/(\sum_{i=1}^{K} \frac{1}{\rho_i c_i})$, we can equivalently set $U(x) = -1/x$. It then follows that $U'^{-1}(y) = -1/y^2$ and (38) becomes

$$\rho_i = \frac{1}{b_i} \left[ \frac{\nu b_i}{c_i} \right]_+ = \frac{\nu}{b_i c_i} = \frac{Q}{b_i c_i} \quad (41)$$

To illustrate the impact of the utility optimization, we consider a surface of size $L = H = 10x$ and drop $K = 5$ users with uniformly distributed azimuth angles $\varphi_i \in [-\pi/2, \pi/2]$ and $\omega_i = 0$. The reference SNR is 0 dB. Fig. [5] shows the CDF of the SE achieved by an arbitrary user when using ZF precoding and the three utilities exemplified above.

Interestingly, the three utilities give similar CDF curves but there are anyway large variations in SE for different user drops. The reason is that the interference is low except when two users happen to get roughly the same angle.

V. CONCLUSION

Spatial interference suppression is important to achieve high spectral efficiency when using an LIS. ZF precoding outperforms MR for practical surface sizes, but we proved that the difference vanishes asymptotically. When using ZF, the power allocation can be efficiently optimized for different utility functions. Although an ideal LIS is a continuous surface, its beam pattern is closely approximated when using discrete antennas of size $\lambda/4 \times \lambda/4$. While this paper considered the far-field, the near-field should be analyzed in future work.

REFERENCES

[1] S. Hu, F. Rusek, and O. Edfors, “Beyond massive MIMO: The potential of data transmission with large intelligent surfaces,” IEEE Trans. Signal Process., vol. 66, no. 10, pp. 2746–2758, 2018.

[2] ——, “Beyond massive MIMO: The potential of positioning with large intelligent surfaces,” IEEE Trans. Signal Process., vol. 66, no. 7, pp. 1761–1774, 2018.

[3] D. Tse and P. Viswanath, Fundamentals of wireless communications. Cambridge University Press, 2005.

[4] E. Björnson, J. Hoydis, and L. Sanguinetti, “Massive MIMO networks: Spectral, energy, and hardware efficiency,” Foundations and Trends® in Signal Processing, vol. 11, no. 3-4, pp. 154–655, 2017.

[5] C. Masouros and M. Matthaiou, “Space-constrained massive MIMO: Hitting the wall of favorable propagation,” IEEE Commun. Lett., vol. 19, no. 5, pp. 771–774, 2015.

[6] E. Björnson, J. Hoydis, and L. Sanguinetti, “Massive MIMO has unlimited capacity,” IEEE Trans. Wireless Commun., vol. 17, no. 1, pp. 574–590, 2018.

[7] S. Hu, F. Rusek, and O. Edfors, “Capacity degradation with modeling hardware impairment in large intelligent surfaces,” in IEEE Global Communications Conference (GLOBECOM), 2018.

[8] E. Björnson, M. Bengtsson, and B. Ottersten, “Optimal multiuser transmit beamforming: A difficult problem with a simple solution structure,” IEEE Signal Process. Mag., vol. 31, no. 4, pp. 142–148, 2014.

[9] E. Björnson and E. Jorswieck, “Optimal resource allocation in coordinated multi-cell systems,” Foundations and Trends® in Communications and Information Theory, vol. 9, no. 2-3, pp. 113–381, 2013.

[10] H. T. Friis, “A note on a simple transmission formula,” IRE, vol. 34, no. 5, pp. 254–256, 1946.

[11] E. Björnson and L. Sanguinetti, “Demystifying the power scaling law of intelligent reflecting surfaces and metasurfaces,” in Proc. IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), 2019.

[12] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, “Scaling up MIMO: Opportunities and challenges with very large arrays,” IEEE Signal Process. Mag., vol. 30, no. 1, pp. 40–60, 2013.