Water-filling with Channel Correlation

Leif Hanlen

Abstract—We show a simple method for increasing on-average capacity of correlated MIMO channels, given limited channel knowledge at the transmitter. We further show that under certain channel constraints, this capacity approaches the capacity gained by water-filling on the complete channel.

We give results which determine the benefit of applying a de-correlating, or whitening, matrix to the transmit signals in terms of the asymptotic growth of the MIMO channel.

Index Terms—MIMO Channel Capacity, Correlated MIMO Channels, Decorrelation, Water-filling

I. INTRODUCTION

Following the remarkable growth predictions for MIMO channels of [1–3], several authors have addressed various difficulties arising from the assumption of receiver channel knowledge and i.i.d. channel behaviour. Spatial correlation has been shown to occur due to placement of transmitters and receivers, or even due to the arrangement of scattering bodies. So-called “pin-hole” [4] or “key-hole” [5] channels have been presented with severe correlation. These channels have the unfortunate property that even with large element separation (> 2λ) the channel remains correlated due to restricted scatter. The work of [6] and [7] discussed significant ramifications on the dense placement of antenna elements, concluding that extreme correlation could occur, independent of the amount of scatter present.

Recent work [8, 9] has shown that the growth in channel capacity for a multiple-input, multiple-output channel is linear in proportion to the number of receive elements, given a large number of transmitters. This is only conditional upon having independent noise at the receiver and does not require an i.i.d. channel. The rate of capacity growth is, however, a function of the correlation of the MIMO channel: although all MIMO channels exhibit a linear growth, severely correlated channels have much lower growth rates for the same numbers of receivers.

It seems sensible to assume that MIMO channels are stuck with correlation in a physical setting. Moreover, it appears that once introduced to the channel, correlation cannot be overcome by subsequent “downstream” alterations. We ask then, might intelligent signal processing techniques allow us to avoid the difficulties of correlated channels?

A common solution for poor channel performance is to provide the transmitter with feedback on the channel. Given knowledge of the transfer function, the ideal transmit strategy is to use some form of water filling on the equivalent set of parallel channels, to optimize power allocation [10, Theorem 7.5.1]. While it is theoretically feasible for the receiver to estimate a given MIMO channel, it is a far more difficult task to subsequently provide channel details to the transmitter. This has been addressed by several authors [11, 12] and a common conclusion is that although it is theoretically beneficial to transmit channel estimation data to the transmitter, the real benefits are usually small.

This is for the following reasons:

• The amount of data required by the transmitter is $O(m^2)$ where $m = \min\{r, t\}$ and $r$ is the number of receivers and $t$ is the number of transmitters.
• A (fast) random fading or rotating channel may change after a small number of symbol periods, requiring retransmission of channel data.

Some authors [13] have recently proposed providing details of the covariance matrix to the transmitter, rather than using a complete water-filling solution. A primary advantage of such an approach is that the correlation matrix of a MIMO channel is likely to change at a much lower rate than the channel itself.

The purpose of this paper is to outline a conditioning for the transmitter, given an estimate of the correlation of the (random) transfer matrix $\mathbf{H}$. We show that knowledge of the correlation matrix provides a large amount of information to the transmitter, even where the transmitter only knows a sub-component of the correlation matrix.

II. MODEL

We shall consider a point-to-point communication link with $t$ transmit antennas and $r$ receive antennas. Throughout the paper we shall refer to $m = \min\{r, t\}$ and $n = \max\{r, t\}$.

At each symbol interval, the received vector $\mathbf{y}$ depends on the transmitted vector $\mathbf{x} \in \mathbb{C}^r$ according to

$$y = \mathbf{H}x + w$$

(1)

Element $y_j$ is the matched-filter output from antenna $j$, while $x_i$ is the signal transmitted from antenna $i$, with the transmitter given a transmission power limit $P$. The matrix $\mathbf{H} \in \mathbb{C}^{r \times t}$ has elements $H_{ji}$, which are the complex gains between transmit antenna $i$ and receive antenna $j$. The vector $w \in \mathbb{C}^r$ contains i.i.d. circularly symmetric Gaussian noise samples [14, p. 134], $\mathcal{E}[ww^\ast] = \eta^2 I_r$.

The channel matrix $\mathbf{H}$ is chosen at random independently each symbol interval, with entries chosen from a complex Gaussian ensemble with zero mean and a covariance matrix $\Sigma$. In the notation of [15] we have $\mathbf{H} \sim \mathcal{N}_{r \times t} (0, \Sigma \otimes I)$, where $\Sigma$ is $m \times m$.

We shall assume that the channel realization $\mathbf{H}$ is known at the receiver at each symbol period and that $\mathbf{H}$ is unknown at the transmitter although some knowledge of $\Sigma$ is available at the...
transmitter. For the purposes of this analysis we shall assume that \( \Sigma \) is fixed, however, this requirement may be relaxed, such that \( \Sigma \) changes slowly over time.

The capacity of the MIMO channel is given by [1]:

\[
C = \mathcal{E} \left[ \log \det \left( I_m + HQH^* \right) \right]
\]

where \( Q = \mathcal{E} \left[ xx^* \right] \) is the covariance matrix of the transmit signals.

### III. WATER-FILLING

For a transmitter with no channel knowledge,\(^1\) the optimal choice of \( Q \) is

\[
Q = \frac{P}{l} I_l \tag{3}
\]

such that the transmitter sends equal power Gaussian circularly symmetric (white) signals. This is the well known “transmit white and hope for the best” strategy and results in the various analytic capacity formulae for uncorrelated [1, 3] and correlated [8] channels available in the literature. We shall denote the channel capacity for white, equal power transmission as \( C_{eq} \).

For a transmitter which has some knowledge of the channel, the optimal strategy is to artificially correlate the transmit signal, using a covariance matrix \( Q \), in such a way as to maximize capacity. If we consider various transmit strategies, the equal power strategy may considered as a lower bound on channel capacity with respect to transmitter knowledge of the channel. For complete knowledge – the transmitter knows \( H \) before it sends signal \( x \) – the optimal approach is the water filling algorithm. Water-filling provides an upper bound on capacity vs transmitter knowledge. For comparison, we present the water-filling algorithm below:

#### Algorithm 1 (Water filling for estimate \( \hat{H} \))

For each instantiation \( \hat{H} \):

1) Receiver calculates \( U \) and \( L \) for \( \hat{H} \) using singular value decomposition \( \hat{HH}^* = \hat{U}L^* \hat{U}^* \), with \( L = \text{diag}(\lambda_1, \ldots, \lambda_m) \)

2) \( Q = ULU^* \) given to transmitter to adjust next symbol \( x \), where \( \hat{D} \) is diagonal and optimizes

\[
\hat{D}_i = \max \left( \mu - \lambda_i^{-1}, 0 \right) \tag{4}
\]

subject to

\[
P = \sum_{i=1}^{m} \max \left( \mu - \lambda_i^{-1}, 0 \right) \tag{5}
\]

where \( \mu \) is the parameter used for optimizing (4).

3) Transmitter sends \( \hat{x} = Q^{1/2} x \) where \( x \) is white.

Algorithm 1 must run at least as fast as the fading rate of the channel. There are modifications to this algorithm which attempt to mitigate estimation errors in \( \hat{H} \) [12] however, we shall focus on the simple algorithm.

\(^1\) No knowledge meaning that the transmitter does not have any information regarding the channel characteristics. We shall compare this with partial channel knowledge, where although the transmitter doesn’t know the particular instantiation \( H \) it does know something about \( \Sigma \).

We shall make the assumption that the receiver keeps an average of the channel estimates \( \hat{H}^n \). Consequently, the amount of work (per estimate) performed at the receiver is the same whether it estimates \( \Sigma \) or \( H \) itself. In the remainder of this paper we shall refer to “water-filling on \( M \)” where \( M \) is a given matrix.

In each case, we may interpret this as using the above algorithm, with \( H \) replaced with \( M \).

We may write a correlated random matrix \( H \sim N_{r,t} \left( 0, \Sigma \otimes I \right) \) as:

\[
H = \Sigma^{1/2} X \Sigma^{1/2} \tag{6}
\]

where \( X \) is the equivalent i.i.d. channel. This immediately suggests an approach for removing channel correlation: measure the correlation \( \Sigma \) at the receiver, and apply \( \hat{x} = \Sigma^{-1/2} x \) at the transmitter. However, such an approach is not feasible, as it contradicts the power limit for the transmitter: highly correlated channels result in large values of \( \Sigma^{1/2} \). In the same manner that waterfilling on \( H \) is the optimal response given a knowledge of \( H \), waterfilling on \( X \) can be shown to be optimal for capacity given only \( \Sigma \) as the knowledge of the channel.

#### Lemma 1.

Given a correlated Gaussian channel \( H \sim N_{r,t} \left( 0, \Sigma \otimes I_n \right) \), with covariance matrix \( \Sigma \), where the transmitter has full knowledge of \( \Sigma \), but does not know \( H \), the input entropy is maximised by waterflling on \( \pi \epsilon \Delta \), where \( \Delta = \text{diag}(\sigma_1, \ldots, \sigma_m) \) and \( \sigma_i \) is an eigenvalue of \( \Sigma \).

**Proof:**

Given \( H \sim N_{r,t} \left( 0, \Sigma \otimes I_n \right) \) we may write:

\[
C = \mathcal{E} \left[ \log \det \left( I_m + \Sigma^{1/2} XQX\Sigma^{1/2} \right) \right] \tag{7}
\]

where \( X \sim N_{r,t} \left( 0, I_m \otimes I_n \right) \), that is, \( X \) is i.i.d with identity covariance matrix. Since \( \Sigma \) is Hermitian, we have:

\[
\Sigma = U \Delta U^* \tag{8}
\]

where \( U \) is unitary and \( \Delta \) is diagonal. With the transformation \( M = \Sigma^{1/2} X \Sigma^{-1/2} \) which has unitary Jacobian, we may write:

\[
C = \mathcal{E} \left[ \log \det \left( I_m + M \Delta^{1/2} U^* QU \Delta^{1/2} M \right) \right] \tag{9}
\]

where we have noted that unitary transforms do not change the distribution \( H \). Since \( X \) is i.i.d. then so is \( M \). From [1, Lemma 2] we have for circularly symmetric Gaussian input \( x \), to the channel \( M \), the entropy is maximum, with

\[
H(x) = \log \det \left( \pi \epsilon \Delta^{1/2} U^* QU \Delta^{1/2} \right) \tag{10}
\]

or

\[
H(x) = \log \det \left( \pi \epsilon \Delta \hat{Q} \right) \tag{11}
\]

where \( \hat{Q} \) is diagonal.

The optimal choice of \( \hat{Q} \) will give the maximum possible entropy, with constraint

\[
\sum_{j=1} Q_{jj} = P \tag{12}
\]
However, \( \log \det(A) \) is strictly concave on positive definite Hermitian matrices \( A \) with maximum at \( A = I \) [16, Theorem 7.6.7]. Since \( Q \) (and \( \Delta^{1/2}U^*QU^{1/2} \)) is positive definite, we can see immediately that the choice of \( \hat{Q} \) which maximises the entropy of the channel is \( \alpha \Delta^{-1} \) for \( \alpha > 0 \).

By itself, lemma 1 is not very remarkable: if we attempt to invert the correlation \( \Sigma \) we should maximise the entropy of the channel by approaching an i.i.d. scenario. Further, since water-filling on \( \Sigma \) cannot be better than water-filling on \( H \), one may wonder why we would attempt to use \( \Sigma \) at all.

There are two reasons for water-filling on \( \Sigma \) rather than \( H \):

- Under fast fades \( H \) is unreliable [13]. The estimate \( \hat{\Sigma} \) is far more stable, and can be estimated with error \( O(N^{-2}) \) [17] given \( N \) samples of \( HH^* \).
- For a large number of transmitters \( t \), the particular instantiation of \( HH^* \) is dominated by \( \Sigma \) [8]. As such, we should expect to achieve similar gains by water-filling on \( \Sigma \) as we would by water-filling on \( H \) directly.

### A. Water-filling on \( \Sigma \)

The work of [8] showed that for finite \( r \), as \( r \) becomes large, the MIMO channel becomes dominated by the eigenvalues of the correlation matrix \( \Sigma \). The capacity of the channel – given no channel knowledge at the transmitter – in the limit of large \( t \) is given by:

\[
C_{eq} = \log \det(I_r + P\Sigma)
\]

(13)

where we have used the simplified asymptotic result from [8]. In terms of various water-filling strategies, we may write:

\[
C_{eq} \leq C_{w,\Sigma} \leq C_{w,H}
\]

(14)

where the transmit strategies are

- \( C_{eq} \): equal power, no channel knowledge,
- \( C_{w,\Sigma} \): water-filling given knowledge of \( \Sigma \) and
- \( C_{w,H} \): water-filling given knowledge of \( H \).

For the case of i.i.d. channel \( C_{w,H} = C_{eq} \). In the case of water-filling given \( \Sigma \), we optimize (13) with respect to the transmit signal covariance matrix \( Q \). We may write:

\[
C_{w,\Sigma} = \sup_Q \log \det \left( I_r + P\Sigma^{1/2}Q\Sigma^{1/2} \right)
\]

(15)

where \( Q \) is constrained by \( \text{trace}(Q) = m \). We may immediately see that the optimal choice of \( Q \) diagonalises \( \Sigma \). In addition, (15) with lemma 1 shows that the optimal \( Q \) found by water-filling on \( \Sigma \) is the same as the optimal \( Q \) found by water-filling on \( H \).

### IV. WATER-FILLING BENEFIT

We may now examine the benefit water-filling gives for a correlated matrix. Using the asymptotic approximation above, we shall assume that \( \Sigma = \text{diag}(\sigma_1, \ldots, \sigma_m) \) where \( \sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_m \).

\[
\sigma_i \approx \begin{cases} c \geq 1 & ; i \leq q \\ 1/m & ; i > q \end{cases}
\]

(16)

This assumption requires the transmitter to have an estimate not only of the eigenvalues of the matrix \( \Sigma \), but also the eigenvectors of \( \Sigma \). Without knowing the eigenvectors the transmitter cannot know the “direction” in which to transmit, and consequently any attempt at water-filling without knowing the eigenvectors will actually be detrimental.

We then write the capacity from water-filling (15) as:

\[
C_{w,\Sigma} = \sum_{i=1}^{m} \log \left( 1 + \left( \mu - \frac{1}{\sigma_i} \right)^+ \sigma_i \right)
\]

(17)

\[
= \sum_{i=1}^{k} \log (\mu\sigma_i)
\]

(18)

where \( k \) is the last eigenvalue such that \( \left( \mu - \frac{1}{\sigma_k} \right) \geq 0 \). If \( \mu \) is larger than \( m \), then we may write the simple expression:

\[
C_{w,\Sigma} \approx q \log(\mu) + (m-q) \log(\mu-m)
\]

(19)

where \( \mu = P + m \). Substituting gives:

\[
C_{w,\Sigma} \approx q \log(P) \log(m) + (m-q) \log(P)
\]

(20)

We may see immediately that (20) predicts a trade-off between linear growth, on \( m-q \) and logarithmic growth in \( m \). The linear growth corresponds to the reduced linear growth predicted by [8] while the logarithmic growth is a result of beamforming. Asymptotically, we may see that water-filling on the correlation matrix will not lead to improved linear growth, although it may result in improved capacity when beamforming is optimal.

For \( \frac{1}{c} \leq \mu \leq m \) we have:

\[
C_{w,\Sigma} \approx q \log \left( \frac{Pm}{q} + 1 \right)
\]

(21)

which displays the form shown in Fig. 1.

![Fig. 1. Growth for small power water-filling, \( P = 1 \).](image)

### V. SIMULATION

We have considered various levels of correlation and transmit power for the transmit strategies \( C_{eq}, C_{w,\Sigma} \) and \( C_{w,H} \). For each plot we have measured the capacity \( C \) in nat/s, and assumed unitary noise variance. To compare equivalent correlation matrices we have adopted a trace-rule [8] such that

\[
\text{trace}(\Sigma) = m
\]

(22)

For equal eigenvalues, (22) gives \( \Sigma = I_m \). We may regard the trace-rule as a power constraint on the channel.
We used the ratio:
\[ \epsilon = \frac{\sigma_{\text{max}}}{\sigma_{\text{mean}}} = \frac{\sigma_{\text{max}}}{\sigma_{\text{mean}}} \] (23)
to compare severity of correlation. The second equality in (23) arises from (22), which defines \( \sigma_{\text{mean}} = 1 \). Larger values of \( \epsilon \geq 1 \) correspond to more extreme correlation.

Fig. 2 shows a correlated matrix, with \( \epsilon = 2 \). Fig. 2(a) gives the capacity for moderate SNR, while Fig. 2(b) gives the capacity for low SNR. For low SNR, \( P = 0.1 \) we find that the \( C_{w,\Sigma} \) strategy is much closer to the optimal water-filling \( C_{w,\Sigma} \) strategy, giving a significant improvement over the equal power strategy. As SNR increases, this benefit diminishes until the \( C_{w,\Sigma} \) strategy does not improve upon the equal power choice.

This can be understood as the optimality of beamforming over dispersive transmission. For very low SNR levels, the optimal choice of transmit strategy is to perform beamforming on the (few) eigenvectors\(^3\) with reasonable strength. These few eigenvectors are well related to the eigenvectors of the correlation matrix: since the correlation matrix dominates the transfer matrix. Conversely, the equal power strategy wastes the majority of its energy, transmitting on a large number of poor eigenvectors.

As the SNR increases, however, smaller eigenvalues (and the corresponding eigenvectors) become more significant. These eigenvectors are not well estimated from the correlation matrix: in fact, it is common practice in statistics to assume the small eigenvalues are spherically distributed noise components [17]. As such, the correlation water-filling approach begins to waste energy by preferentially choosing the few large eigenvalues over the smaller ones.

It is important to realize, however, that knowledge of the singular values of \( \Sigma \) is less critical than knowledge of the eigenvectors. This is shown in Fig. 3, where we have compared \( C_{w,\Sigma} \) with \( C_{w,\Sigma_{1/2}} \). The strategy of \( C_{w,\Sigma_{1/2}} \) performed water-filling on \( \Sigma_{1/2} \) rather than \( \Sigma \). We can see for this strategy that the capacity is not improved by as great an amount, although there is some benefit. The eigenvectors may be considered as directions for transmission for beamforming. Clearly, if we maintain the optimal directions but do not transmit the ideal amount of

\(^3\) Technically beamforming is where the transmitter uses only one eigenvector. All power is transmit in this particular direction. In the case of multiple (equal strength) eigenvectors, we may interpret the transmit strategy as “beamforming” in all directions at once: perhaps using multiple beams. Equally, we may interpret the strategy as a reduced MIMO arrangement.
power in each, we should still see some improvement in channel capacity in all but the most pathological cases. Water-filling, without the correct eigenvectors results in near-zero capacity on average. This can be seen by realizing that for all possible directions (all unitary matrices) very few will give transmission in the appropriate directions. The ratio can be calculated as $P_r(M = U) = V_{O(n)} V_{O(n)}$, where $V_{O(n)}$ is the volume of the Stiefel manifold and $V_{O(n)}$ is the volume of the equivalent manifold over permutation matrices.

In each case, comparing the asymptotic growth in $C_{eq \Sigma}$ with $C_{eq}$ reveals that the rate of growth is unchanged by water-filling on the correlation matrix. There is, however, a (possibly significant) capacity offset from the equal power strategy, which is due to beamforming on a small number of strong eigenvectors.

This should not be interpreted as a negative result for water-filling: the basic result is that water-filling does improve capacity, however, when we use water-filling on the correlation matrix only, we find that the rate of growth is not improved. This adds weight to the use of asymptotic results based on zero knowledge channel assumptions.

VI. CONCLUSION

Correlated MIMO channels are tantalizing in that they offer many, but not all, of the benefits of i.i.d. MIMO channels. We have examined the possibility of water-filling using the correlation matrix rather than the channel itself. This has the advantage that channel correlation does not change as fast as the channel does.

We have shown that a significant improvement can be made for MIMO channels using this water-filling strategy. We have also shown, however, that the improvement appears in the form of an offset to the linear growth of the correlated channel. We have found that given a correlated channel with capacity $C = \nu \times m$ for some $\nu$, the water-filling strategy results in a channel with capacity $C = \nu \times m + \log(m)$.

This result suggests that while water-filling may offer some hope in improving channel capacity, the main limit on capacity growth remains the correlation of the channel.

REFERENCES


