POWER ALLOCATION IN MULTI-ANTENNA COMMUNICATION WITH STATISTICAL CHANNEL INFORMATION AT THE TRANSMITTER

Antonia M. Tulino[1], Angel Lozano[2], Sergio Verdú[3]

[1] Università di Napoli "Federico II", Italy, atulino@ee.princeton.edu
[2] Bell Labs (Lucent Technologies), USA, aloz@lucent.com
[3] Princeton University, USA, verdu@ee.princeton.edu

Abstract - We characterize the power allocation that maximizes the rate per unit bandwidth supported with arbitrary reliability over single-user multi-antenna channels known instantaneously by the receiver and in distribution by the transmitter. The characterization is valid for arbitrary channels and numbers of antennas. Although, in general, it leads to a fixed-point solution, at low and high signal-to-noise it provides explicit allocations. For arbitrary signal-to-noise ratios, we present an iterative algorithm that exhibits remarkable properties: robustness, rapid convergence and universal applicability. Further, when applied to the proper set of signalling eigenvectors, the algorithm converges to the power allocation that attains capacity.

I. Motivation

While, in most instances of wireless communication, the receiver can accurately track the instantaneous state of the channel, the transmitter is often unable to perform such tracking. Statistical information about the channel, on the other hand, is virtually always accessible to the transmitter since the periods over which a fading process is basically stationary are several orders of magnitude larger than the duration of the fades. As a result, the most typical operating regime in mobile systems is that in which (*i*) the receiver knows the channel instantaneously, and (*ii*) the transmitter has only access to its distribution. In such regime, which constitutes the focus of this paper, the transmitted signal cannot be tailored to the state of the channel, but only to its distribution.

Let us consider a single-user channel with n_{T} transmit and n_{R} receive antennas and let us denote the transmitted signal vector by $\mathbf{x}$. In the presence of Gaussian noise, achieving capacity requires that $\mathbf{x}$ be zero-mean and Gaussian. The key quantity to determine is thus its spatial covariance, given by

$$\mathbf{\Phi} \triangleq \frac{E[\mathbf{x}\mathbf{x}^{\dagger}]}{\frac{1}{n_{\mathrm{T}}}E[\|\mathbf{x}\|^2]}$$

conveniently normalized by its energy per dimension so that $E[\mathrm{Tr}\{\mathbf{\Phi}\}]{=}n_{\mathrm{T}}$.

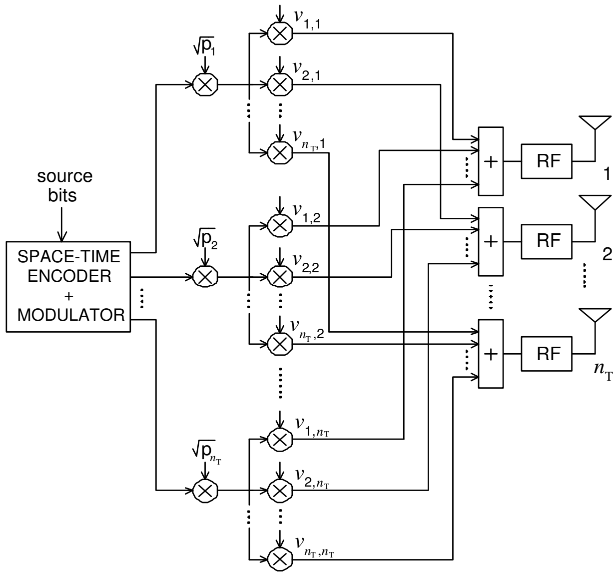

Fig. 1
Architecture generating a signal with covariance $\mathbf{\Phi}{=}\mathbf{V}\mathbf{P}\mathbf{V}^{\dagger}$ where $\mathbf{V}{=}[v_{i,j}]$ is unitary and $\mathbf{P}{=}\mathrm{diag}\{p_1, p_2, \ldots, p_{n_{\mathrm{T}}}\}$.

In order to evidence the two fundamental levels of control that can be exercised on $\mathbf{x}$ through $\mathbf{\Phi}$, it is useful to decompose the latter as $\mathbf{\Phi}{=}\mathbf{V}\mathbf{P}\mathbf{V}^{\dagger}$ identifying the eigenvectors of $\mathbf{\Phi}$ with the columns of the unitary matrix $\mathbf{V}$ and its eigenvalues with the diagonal entries of $\mathbf{P}{=}\mathrm{diag}\{p_1, p_2, \ldots, p_{n_{\mathrm{T}}}\}$. Both the eigenvectors and the eigenvalues have immediate engineering meaning: the former indicate the *directions* (in vector space) on which signalling takes place while the latter signify the normalized *powers* allocated onto each such direction. This is illustrated in Fig. 1, which depicts a transmit architecture that generates a signal with arbitrary spatial covariance. For each channel use, a space-time encoder outputs a set of n_{T} parallel symbols, each of which is assigned a certain power (which may be zero) and rotated into a certain direction before being simultaneously radiated out of the n_{T} antennas.

Unlike in the regime where the transmitter can track the channel state instantaneously, where the capacity-achieving forms for $\mathbf{V}$ and $\mathbf{P}$ are well established [1], for our regime of interest:

- The signalling eigenvectors on which capacity is attained are known only for certain classes of channels [1]–[4].
- The power allocation had, to date, been left mostly to numerical optimization.

In this paper, we show how this power allocation can be addressed analytically. Specifically,

- We present necessary and sufficient conditions characterizing the powers that maximize the rate supported on any set of signalling eigenvectors. In some limiting scenarios, at low- and high-SNR, these conditions lead to explicit solutions.
- For arbitrary SNR, an iterative algorithm is unveiled to solve for the powers.
- When applied to the capacity-achieving signalling eigenvectors, this algorithm finds the power allocation that achieves capacity.

II. Problem Formulation

The baseband complex model we consider is

$$\mathbf{y} = \sqrt{g}\,\mathbf{H}\,\mathbf{x} + \mathbf{n}$$

where $\mathbf{y}$ is the received vector while $\mathbf{n}$ is a white Gaussian noise vector. The channel is represented by the $(n_{\mathrm{R}} \times n_{\mathrm{T}})$ random matrix $\sqrt{g}\,\mathbf{H}$ where the scalar g is such that

$$E[\mathrm{Tr}\{\mathbf{H}\mathbf{H}^{\dagger}\}] = n_{\mathrm{R}} n_{\mathrm{T}}.$$

The average rate per unit bandwidth (in b/s/Hz) that can be supported with arbitrary reliability is

$$I(\mathsf{SNR}) = E\left[\log_2 \det\left(\mathbf{I} + \frac{\mathsf{SNR}}{n_{\mathrm{T}}}\,\mathbf{H}\mathbf{V}\mathbf{P}\mathbf{V}^{\dagger}\mathbf{H}^{\dagger}\right)\right] \quad (1)$$

with $\mathrm{Tr}\{\mathbf{P}\} = n_{\mathrm{T}}$ and with

$$\mathsf{SNR} \triangleq g\,\frac{E[\|\mathbf{x}\|^2]}{\frac{1}{n_{\mathrm{R}}}E[\|\mathbf{n}\|^2]}$$

Our objective is to characterize the power allocation, $\mathbf{P}$, that maximizes (1) for any channel distribution given any set of signalling eigenvectors, $\mathbf{V}$.

III. Rate-Maximizing Power Allocation

Define a rotated channel, $\hat{\mathbf{H}} = \mathbf{H}\mathbf{V}$, whose j-th column is denoted by $\hat{\mathbf{h}}_j$. In order to maximize $I(\mathsf{SNR})$, $\mathbf{P}$ must satisfy the following set—derived in the Appendix—of necessary and sufficient conditions:

$$E\left[\mathrm{Tr}\left\{\left(\mathbf{I} + \frac{\mathsf{SNR}}{n_{\mathrm{T}}}\,\hat{\mathbf{H}}\mathbf{P}\hat{\mathbf{H}}^{\dagger}\right)^{-1}\right\} + \mathsf{SNR}\,\hat{\mathbf{h}}_j^{\dagger}\left(\mathbf{I} + \frac{\mathsf{SNR}}{n_{\mathrm{T}}}\,\hat{\mathbf{H}}\mathbf{P}\hat{\mathbf{H}}^{\dagger}\right)^{-1}\hat{\mathbf{h}}_j\right] \begin{array}{ll} = n_{\mathrm{R}} & \text{if } p_j > 0 \\ \leq n_{\mathrm{R}} & \text{if } p_j = 0 \end{array} \quad (2)$$

A unique set of powers exists that satisfies (2). Since the corresponding parallel signalling channels are not orthogonal, this solution does not correspond to a waterfill on any statistical measure of the channel.[1] The following limiting behaviors can be observed:

- For $\mathsf{SNR} \to 0$, power should be allocated only to the signalling eigenvector whose index corresponds with the maximal diagonal entry of $E[\hat{\mathbf{H}}^{\dagger}\hat{\mathbf{H}}]$. If the multiplicity of such maximal value is plural, equal power should be assigned to the corresponding signalling eigenvectors (see [6]).
- For $\mathsf{SNR} \to \infty$, power should be allocated only to the eigenvectors whose indices correspond with nonzero diagonal entries of $E[\hat{\mathbf{H}}^{\dagger}\hat{\mathbf{H}}]$. If the number of nonzero diagonal entries of $E[\hat{\mathbf{H}}\hat{\mathbf{H}}^{\dagger}]$ equals or exceeds the number of nonzero diagonal entries of $E[\hat{\mathbf{H}}^{\dagger}\hat{\mathbf{H}}]$, such powers should be further equal.

Beyond these asymptotes, the set of powers satisfying (2) cannot, in general, be found explicitly. Rather than leaving them to be numerically optimized, though, in the next section we derive—directly from (2)—an iterative power allocation algorithm.

IV. Iterative Algorithm

In order to formulate this algorithm, it is useful to introduce, as an auxiliary quantity, the mean-square error on the linear MMSE (minimum mean-square error) estimation of the signal transmitted along the j-th signalling eigenvector. Defining

$$\mathbf{B}_j \triangleq \left(\mathbf{I} + \frac{\mathsf{SNR}}{n_{\mathrm{T}}}\hat{\mathbf{H}}_j\mathbf{P}_j\hat{\mathbf{H}}_j^{\dagger}\right)^{-1} \quad (3)$$

where $\hat{\mathbf{H}}_j$ indicates the matrix obtained by removing from $\hat{\mathbf{H}}$ the j-th column whereas $\mathbf{P}_j$ indicates the diagonal matrix obtained by removing from $\mathbf{P}$ the (j,j)-th diagonal entry, such mean-square error is [7]

$$\mathsf{MSE}_j = \frac{1}{1 + p_j\frac{\mathsf{SNR}}{n_{\mathrm{T}}}\hat{\mathbf{h}}_j^{\dagger}\mathbf{B}_j\hat{\mathbf{h}}_j} \quad (4)$$

The useful signal power recovered along the j-th eigenvector is $1 - \mathsf{MSE}_j$ and thus the corresponding signal-to-interference-and-noise ratio equals $\frac{1}{\mathsf{MSE}_j} - 1$.

The expectation of (4) with respect to $\hat{\mathbf{H}}$, in turn, is denoted by

$$\overline{\mathsf{MSE}}_j \triangleq E[\mathsf{MSE}_j] \quad (5)$$

and it can be verified that

$$\mathrm{Tr}\left\{\left(\mathbf{I} + \frac{\mathsf{SNR}}{n_{\mathrm{T}}}\hat{\mathbf{H}}\mathbf{P}\hat{\mathbf{H}}^{\dagger}\right)^{-1}\right\} = \sum_{\ell=1}^{n_{\mathrm{T}}}\mathsf{MSE}_{\ell} + n_{\mathrm{R}} - n_{\mathrm{T}} \quad (6)$$

Using (3)–(6), the conditions in (2) can be rewritten as

$$p_j = \frac{1 - \overline{\mathsf{MSE}}_j}{\sum_{\ell=1}^{n_{\mathrm{T}}}(1 - \overline{\mathsf{MSE}}_{\ell})}\,n_{\mathrm{T}} \qquad j = 1, \ldots, n_{\mathrm{T}} \quad (7)$$

[1]This is in contrast with the regime where $\mathbf{H}$ is known by the transmitter, in which case parallel orthogonal channels can be created and the power allocation does reduce to a waterfill [1], [5].

revealing that, in order to maximize $I(\text{SNR})$, the value taken by each power must strike a careful ratio between the average signal power recovered (by an MMSE receiver) from the corresponding eigenvector and the sum of the average signal powers recovered from all the eigenvectors. Hence, those signalling eigenvectors from which a relatively strong average signal power can be recovered should be allocated a larger share of the power budget and viceversa.

The conditions in (7) constitute a set of coupled implicit equations that begets an iterative approach. To accommodate the iterative nature of the resulting algorithm, we shall use $(\cdot)^{(k)}$ to index the succession of values taken by each of the quantities involved.

Algorithm 1 *Initialize* $\mathbf{P}^{(0)}$ *to best available guess with* $\text{Tr}\{\mathbf{P}^{(0)}\}{>}0$. *(If no prior information, set* $\mathbf{P}^{(0)}{=}\mathbf{I}$*). Iterate as follows:*

Step I. *Calculate*

$$p_j^{(k+1)} = \frac{1-\overline{\text{MSE}}_j^{(k)}}{\sum_{\ell=1}^{n_\text{T}}(1-\overline{\text{MSE}}_\ell^{(k)})}\, n_\text{T} \qquad \forall j$$

Step II. *Declare new power allocation*

$$\mathbf{P}^{(k+1)} = \text{diag}\left\{p_1^{(k+1)}, p_2^{(k+1)}, \ldots, p_{n_\text{T}}^{(k+1)}\right\}.$$

Notice that Step I sets $\text{Tr}\{\mathbf{P}^{(k)}\}{=}n_\text{T}$ for $k{>}0$ even if $\text{Tr}\{\mathbf{P}^{(0)}\}{\neq}n_\text{T}$. Hence, the total transmitted power is held at the correct value throughout the iterations as long as the initial powers are not identically zero. Note also that, if a particular power is initialized to zero, it remains at zero indefinitely. Thus, except possibly for those known to be zero, the initial powers in $\mathbf{P}^{(0)}$ should be strictly positive.

As illustrated next through a sequence of examples, the iterations converge rapidly to the sought fixed-point set of powers. This might render the algorithm suitable for implementation in time-varying environments, where the channel distribution itself is subject to slow macroscopic variations that require tracking.

V. Examples

It is chief to realize that the functioning of the algorithm hinges on the availability of *average* mean-square errors, which have to be computed by processing a series of instantaneous observations. Typically, the coherence distance of the channel distribution ranges between a few meters and a few tens of meters. At frequencies of $2-5$ GHz, where the fade duration is on the order of a few cm, this coherence distance may expose no more than a few hundred independent channel realizations whereby those averages must be extracted. Recognizing this fundamental constraint, in the examples that follow the expectations are computed as averages of only 100 independent realizations.

A. Correlated Rayleigh-faded Channel

Consider the channel

$$\mathbf{H} = \mathbf{\Theta}_\text{R}^{1/2}\mathbf{W}\mathbf{\Theta}_\text{T}^{1/2} \tag{8}$$

where the entries of $\mathbf{W}$ are IID (independent identically distributed) zero-mean unit-variance complex Gaussian while $\mathbf{\Theta}_\text{R}$ and $\mathbf{\Theta}_\text{T}$ are $(n_\text{R}{\times}n_\text{R})$ and $(n_\text{T}{\times}n_\text{T})$ correlation matrices whose entries indicate, respectively, the correlation between receive antennas and between transmit antennas.

On such channel, achieving capacity requires that the signaling eigenvectors equal the eigenvectors of $\mathbf{\Theta}_\text{T}$ [2]. Our first example illustrates how the corresponding capacity-achieving power allocation is determined.

Example 1 *Let* $n_\text{T}{=}3$ *on a uniform linear array with* 1*-wavelength antenna spacing. If the channel exhibits a broadside (truncated) Gaussian power azimuth spectrum with* 2° *root-mean-square spread, the transmit correlations are* $(\mathbf{\Theta}_\text{T})_{i,j}{\approx}e^{-0.05\,(i-j)^2}$. *Further consider* $n_\text{R}{=}4$ *uncorrelated receive antennas, i.e.,* $\mathbf{\Theta}_\text{R}{=}\mathbf{I}$. *Signalling over the eigenvectors of* $\mathbf{\Theta}_\text{T}$*, the convergence at* SNR$=-3$ dB *and* SNR$=5$ dB *is depicted in Fig. 2.*

The anticipated low-SNR behavior is manifest at -3 dB, where a single eigenvector is allocated the entire power budget.

If, instead of signalling over the eigenvectors of $\mathbf{\Theta}_\text{T}$ as in Example 1, independent signals are radiated from each transmit antenna, the power allocation is noticeably different.

Example 2 *Recall the conditions of Example 1. With* $\mathbf{V}{=}\mathbf{I}$*, the convergence at* SNR$=5$ dB *is shown in Fig. 3.*

Whereas the capacity at SNR$=5$ dB turns out to be 5.13 b/s/Hz, with $\mathbf{V}{=}\mathbf{I}$ only 4.36 b/s/Hz can be sustained.

B. Ricean Channel

Consider now the channel

$$\mathbf{H} = \sqrt{\tfrac{K}{K+1}}\bar{\mathbf{H}} + \sqrt{\tfrac{1}{K+1}}\mathbf{W}$$

where $\bar{\mathbf{H}}$ is deterministic while $\mathbf{W}$ is as in (8). This is a Ricean channel with factor K, for which the signalling eigenvectors on which capacity is achieved are known to coincide with those of $\bar{\mathbf{H}}^\dagger\bar{\mathbf{H}}$ [3], [4].

Example 3 *Let* $n_\text{T}{=}3$ *and* $n_\text{R}{=}2$ *with*

$$\mathbf{H} = \tfrac{1}{\sqrt{2}}\bar{\mathbf{H}} + \tfrac{1}{\sqrt{2}}\mathbf{W}$$

where the entries of $\bar{\mathbf{H}}$ *equal* 1 *and the Ricean factor is* 0 *dB. Signalling on the eigenvectors of* $\bar{\mathbf{H}}^\dagger\bar{\mathbf{H}}$*, the convergence of the transmit powers to their capacity-achieving values at* SNR$=5$ dB *is portrayed in Fig. 4.*

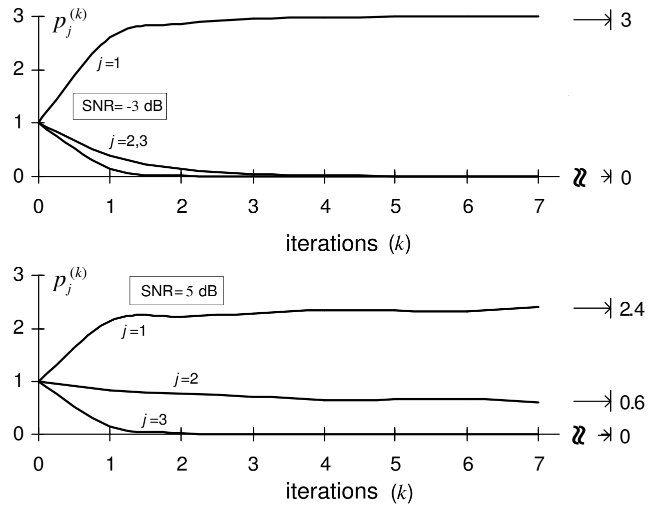

Fig. 2
$\mathbf{P}^{(k)}$ for $k{=}1,\ldots,7$ in Example 1 with $\mathbf{P}^{(0)}{=}\mathbf{I}$ and with $\mathbf{V}$ given by the eigenvectors of $\mathbf{\Theta}_{\mathrm{T}}$. The capacity-achieving powers (obtained numerically) are $\mathbf{P}{=}\mathrm{diag}\{3,0,0\}$ at SNR$=-3$ dB and $\mathbf{P}{=}\mathrm{diag}\{2.4,0.6,0\}$ at SNR$=5$ dB.

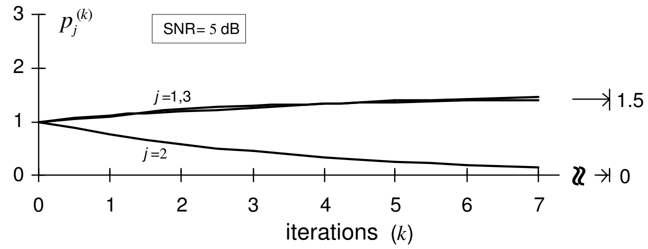

Fig. 3
$\mathbf{P}^{(k)}$ for $k{=}1,\ldots,7$ in Example 2 with $\mathbf{P}^{(0)}{=}\mathbf{I}$ and $\mathbf{V}{=}\mathbf{I}$. Rate-maximizing powers (numerically): $\mathbf{P}{=}\mathrm{diag}\{1.5,0,1.5\}$.

As in the Rayleigh-fading case, it is interesting to evaluate the power allocation under independent transmissions from each antenna, a scenario that fits many space-time encoding techniques.

Example 4 *Recall Example 3. For* $\mathbf{V}{=}\mathbf{I}$*, it is verified that the rate is maximized with* $\mathbf{P}{=}\mathbf{I}$*. The algorithm convergence at* SNR$=5$ dB*, shown in Fig. 5, confirms this allocation.*

The range of variability around the solution relates directly to the number of independent channel realizations available for the estimation of the average mean-square errors. With only 100 realizations, the powers remain remarkably tight around their ideal values.

C. Keyhole Channel

A keyhole channel is modelled as $\mathbf{H}{=}\mathbf{c}_{\mathrm{R}}\mathbf{c}_{\mathrm{T}}^{\dagger}$ where $\mathbf{c}_{\mathrm{R}}$ and $\mathbf{c}_{\mathrm{T}}$ are column vectors with IID zero-mean com-

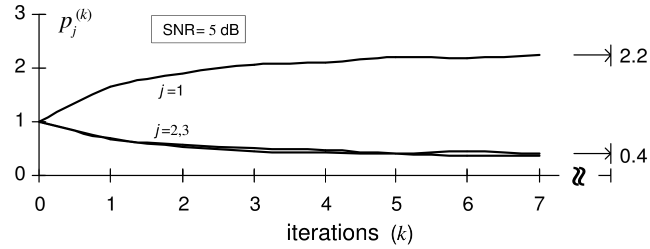

Fig. 4
$\mathbf{P}^{(k)}$ for $k{=}1,\ldots,7$ in Example 3 with $\mathbf{P}^{(0)}{=}\mathbf{I}$ and with $\mathbf{V}$ given by the eigenvectors of $\bar{\mathbf{H}}^{\dagger}\bar{\mathbf{H}}$. The capacity-achieving powers (obtained numerically) are $\mathbf{P}{=}\mathrm{diag}\{2.2,0.4,0.4\}$.

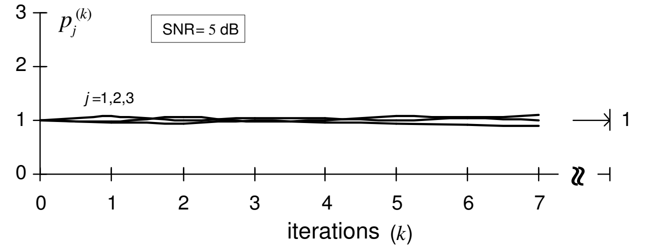

Fig. 5
$\mathbf{P}^{(k)}$ for $k{=}1,\ldots,7$ in Example 4 with $\mathbf{P}^{(0)}{=}\mathbf{I}$ and $\mathbf{V}{=}\mathbf{I}$. The rate-maximizing powers are $\mathbf{P}{=}\mathbf{I}$.

plex Gaussian entries [8]. Its capacity-achieving power allocation is uniform, i.e., $\mathbf{P}{=}\mathbf{I}$ with $\mathbf{V}$ immaterial.

Example 5 *Consider a keyhole channel with* $n_{\mathrm{T}}{=}n_{\mathrm{R}}{=}2$*. Signalling with* $\mathbf{V}{=}\mathbf{I}$*, the convergence of the transmit powers to* $\mathbf{P}{=}\mathbf{I}$ *at* SNR$=8$ *dB is illustrated in Fig. 6. For emphasis, the algorithm is initialized with* $\mathbf{P}^{(0)}{=}\mathrm{diag}\{1.8,0.2\}$*, a strongly non-uniform allocation.*

VI. Conclusions

We have characterized the rate-maximizing power allocation for multi-antenna channels known instantaneously at the receiver while only in distribution at the transmitter. The solution is not—despite several

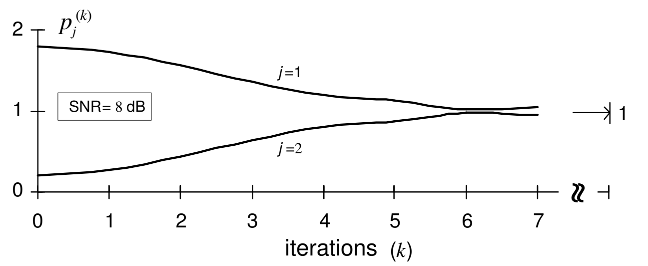

Fig. 6
$\mathbf{P}^{(k)}$ in Example 5 with $\mathbf{P}^{(0)}{=}\mathrm{diag}\{1.8,0.2\}$ and $\mathbf{V}{=}\mathbf{I}$.

claims in the literature—a statistical extension of the waterfill encountered when the transmitter also knows the channel instantaneously. Because of the lack of orthogonality between the parallel transmissions, part of the power radiated on each signalling eigenvector spills as interference onto the other ones and thus the powers are mutually coupled beyond their sum constraint. What we have shown is that, rather than obtained via classic waterfill [5], the fraction of the total available power allocated to each signalling eigenvector should equal the fraction of the average signal power recovered, by an MMSE receiver, from that eigenvector. Although this solution does yield some behaviors that are reminiscent of a waterfill, it is in general quite distinct. Particularly noteworthy are the limiting power allocations at low and high SNR:

- At low SNR, concentrating power on the strongest eigenvector(s) is the rate-maximizing policy.
- At high SNR, where a MMSE receiver behaves in zero-forcing mode, the allocation is sensitive to the relative numbers of transmit and receive antennas. If the number of receivers equals or exceeds the number of transmitters, a zero-forcer can extract the signal from each eigenvector completely removing the interference from the other ones. The powers thus decouple and the resulting policy is a uniform power allocation. With more transmitters than receivers, however, this is no longer the case.

Although in general not explicit, our solution for the powers leads rather straightforwardly to an iterative algorithm, far more alluring than a numerical procedure, whose main features are:

- It is universally applicable. Given an arbitrary channel and a set of eigenvectors, it finds the power allocation that maximizes the rate attained by a Gaussian input signalling thereon.
- Applied to the proper eigenvectors, it finds the capacity-achieving power allocation.
- It requires no step-size or parameter adjustments.
- It is very robust in terms of initial conditions, which are only required to be nonzero.
- It exhibits very rapid convergence, which might render it suitable for adaptive implementation.

Appendix

Let $\mathbf{P}$ be the diagonal matrix that maximizes the strictly concave function (in nats/s/Hz)

$$I(\mathbf{P}_\mathsf{d}) = E\left[\log_e \det\left(\mathbf{I} + \frac{\mathsf{SNR}}{n_\mathrm{T}}\hat{\mathbf{H}}\mathbf{P}_\mathsf{d}\hat{\mathbf{H}}^\dagger\right)\right] \tag{9}$$

over the convex set of diagonal positive-definite matrices $\mathbf{P}_\mathsf{d}$ such that $\mathrm{Tr}\{\mathbf{P}_\mathsf{d}\}{=}n_\mathrm{T}$. Such $\mathbf{P}$ is characterized by a set of Kuhn-Tucker conditions [5]. Concordingly, we impose that the derivative of (9) in the direction from $\mathbf{P}$ to any other matrix $\mathbf{P}_\mathsf{d}$ be negative. Letting

$$\mathbf{P}_\mu = (1-\mu)\mathbf{P} + \mu\mathbf{P}_\mathsf{d}$$

for $0{\le}\mu{\le}1$, the one-side derivative of (9) with respect to μ at $\mu{=}0^+$ is

$$\frac{d}{d\mu}I(\mathbf{P}_\mu) = E\left[\mathrm{Tr}\left\{\left(\mathbf{I} + \frac{\mathsf{SNR}}{n_\mathrm{T}}\hat{\mathbf{H}}\mathbf{P}_\mathsf{d}\hat{\mathbf{H}}^\dagger\right)\left(\mathbf{I} + \frac{\mathsf{SNR}}{n_\mathrm{T}}\hat{\mathbf{H}}\mathbf{P}\hat{\mathbf{H}}^\dagger\right)^{-1} - \mathbf{I}\right\}\right]$$

and, therefore, we impose that

$$E\left[\mathrm{Tr}\left\{\left(\mathbf{I} + \frac{\mathsf{SNR}}{n_\mathrm{T}}\hat{\mathbf{H}}\mathbf{P}_\mathsf{d}\hat{\mathbf{H}}^\dagger\right)\left(\mathbf{I} + \frac{\mathsf{SNR}}{n_\mathrm{T}}\hat{\mathbf{H}}\mathbf{P}\hat{\mathbf{H}}^\dagger\right)^{-1} - \mathbf{I}\right\}\right] \le 0 \tag{10}$$

for every $\mathbf{P}_\mathsf{d}$ in the set. Since (10) is linear on $\mathbf{P}_\mathsf{d}$, it suffices to impose it on the extreme points of the set. Moreover, the line connecting the j-th extreme point ($p_j{=}n_\mathrm{T}$, $p_\ell{=}0$ for $\ell{\neq}j$) with $\mathbf{P}$ can be extended beyond $\mathbf{P}$ if and only if the optimum p_j is strictly positive, in which case the derivative at $\mathbf{P}$ vanishes and (10) is a strict equality. Otherwise, if the optimum p_j is zero, (10) remains an inequality. With these considerations, (2) is readily obtained from (10).

References

[1] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecom*, vol. 10, pp. 585–595, Nov. 1999.

[2] S. A. Jafar, S. Vishwanath, and A. J. Goldsmith, "Channel capacity and beamforming for multiple transmit and receive antennas with covariance feedback," *Proc. IEEE Intern. Conf. on Commun. (ICC'01)*, vol. 7, pp. 2266–2270, 2001.

[3] S. Venkatesan, S. H. Simon, and R. A. Valenzuela, "Capacity of a Gaussian MIMO channel with nonzero mean," *Proc. of IEEE Veh. Tech. Conf. (VTC'03), Orlando, FL*, Oct. 2003.

[4] D. Hosli and A. Lapidoth, "The capacity of a MIMO Ricean channel is monotonic in the singular values of the mean," *Proc. Intern. ITG Conf. on Source & Channel Coding*, pp. 381–385, Jan. 2004.

[5] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, Wiley, 1990.

[6] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. on Inform. Theory*, vol. 48, pp. 1319–1343, June 2002.

[7] S. Verdú, *Multiuser Detection*. Cambridge University Press, 1998.

[8] D. Chizhik, G. J. Foschini, M. J. Gans, and R. A. Valenzuela, "Keyholes, correlations and capacities of multi-element transmit and receive antennas," *IEEE Trans. on Wireless Communic.*, vol. 1, pp. 361–368, Apr. 2002.