IWSSIP 2012, 11-13 April 2012, Vienna, Austria ISBN 978-3-200-02328-4

An Algorithm for Highly Efficient Waterfilling with Guaranteed Convergence

Johannes Gonter and Norbert Goertz

Institute of Telecommunications
Vienna University of Technology
Gußhausstr. 25-29 / 389, 1040 Vienna, Austria
Email: {johannes.gonter, norbert.goertz}@nt.tuwien.ac.at

***Abstract*— This paper derives a novel and highly efficient algorithm for calculating the waterfilling parameter λ (i.e., the water level) for maximization of transmitted information, given a set of channel-power coefficients, and a sum-energy constraint. The resultant fixed-point algorithm calculates λ in at most 9 iterations for 1000 given Rayleigh-distributed channel-power coefficients, each with scale-parameter $\sigma = 0.2$. The proposed fixed-point algorithm guarantees convergence of λ to a unique optimum value, without restricting validity to practically relevant channel-power coefficient distributions.**

I. Introduction

Modern communication systems' number-one optimization objective is efficiency. In some subdomains, this objective translates to energy efficiency or optimum network coverage; for practical wireless systems it may well be spectral efficiency. However, many notions of efficiency reduce to algorithmic efficiency.

This paper derives a novel and highly efficient algorithm for calculating the waterfilling parameter λ (i.e., the water level) in maximization of transmitted information, given a set of channel-power coefficients and a sum-energy constraint. Preservation of generality avoids any restriction of the channel coefficients' statistics. Specifically, the new algorithm does not assume channel ergodicity, since the algorithms' goal is determination of the optimum power-allocation strategy for any given set of channel-power coefficients. Proof of guaranteed algorithm-convergence uses the fixed-point function in a most general fashion, revealing some stunning properties. The algorithm is different from other approaches (for a comprehensive overview, see [1]) in that it introduces a fixed-point approach which circumvents the need for fine-tuning the algorithm according to the specific properties of the dataset it works on. This paper is organized as follows: Section II defines the notion "channel-power coefficients" and states these coefficients' statistical properties for the generation of all this paper's examples. Section III derives the waterfilling solution for maximizing the amount of information transmitted, given a set of channel-power coefficients and a sum-energy constraint. Section IV shows that the waterfilling parameter λ can be efficiently computed by fixed-point iterations. Section IV also finds an equivalent expression for the fixed-point function based on the probability-density function (pdf) and the cumulative distribution function (cdf) of the channel-power coefficients. Exploiting most general statistical properties of the channel-power coefficients proves guaranteed convergence of the fixed-point iterations to the optimum waterfilling-parameter. Section V presents a minimum-complexity Matlab-implementation of the fixed-point algorithm and shows the convergence behavior for the exemplary Rayleigh-distributed channel-power coefficients and for randomly chosen initial values of λ. Section VI concludes the paper.

II. Preliminary Definitions

The channel model is the common block-fading with i.i.d. Gaussian noise. Throughout, this paper only considers a finite number of channel-power coefficients at once, all of which are required to be non-zero. This development therefore defines $\mathcal{S} \doteq \{i : |h(i)|^2 > 0\}$, where $|h(i)|^2$ are the channel-power coefficients upon which the new fixed-point algorithm works. Therefore, $|\mathcal{S}| < \infty$. In each block i, the receive power $P_r(i)$ will equal the transmit power $P(i)$ scaled by the magnitude squared $|h(i)|^2$ of the channel coefficient $h(i)$. Therefore, the channel-power coefficient is

$$P_r(i) = |h(i)|^2 \cdot P(i) \ . \tag{1}$$

Most general validity of the new algorithm considers the probability distribution of the observed channel-power coefficients to be that distribution generated by sampling a non-ergodic channel at different but deterministic time-instants. General presentation of numerical results uses Rayleigh-distributed channel-power coefficients with parameter $\sigma = 0.2$ for the simulations and examples. The proposed schemes are, however, not restricted to any specific coefficient distribution.

III. The Optimization Problem

The capacity $C(i)$ (in bits per channel-use) within block i is given by

$$C(i) = \log_2 \big(1 + \frac{|h(i)|^2 \cdot P(i)}{2\sigma^2}\big) \tag{2}$$

where σ^2 is the variance of the Gaussian receiver noise in each real component of the transmit channel (in-phase and quadrature component), implicitly assuming complex modulation by definition of (2). The total power $P(i)$ is spread equally across the two real channel parts, because this will, for symmetric channels, achieve capacity (e.g., [2]). Each block i contains M channel uses. Then the number of bits which can be transmitted in each block i is at most

$$N = M \sum_{\forall i \in \mathcal{S}} \log_2 \big(1 + \frac{|h(i)|^2 \cdot P(i)}{2\sigma^2}\big) . \tag{3}$$

The algorithm will maximize the number N of bits in (3), given $|h(i)|^2$ for all $i \in \mathcal{S}$. Of course large power can achieve an arbitrarily large N, so the problem adds an energy constraint. The total energy[1] used for the transmission of the information bits by (compare [3]) is

$$E = \sum_{\forall i \in \mathcal{S}} MP(i) = M \sum_{\forall i \in \mathcal{S}} P(i) . \tag{4}$$

Limitation of the value of the total energy according to

$$0 \leq E = M \sum_{\forall i \in \mathcal{S}} P(i) \leq E_0 , \tag{5}$$

defines a constrained optimization problem. The corresponding functional L with the Lagrange multiplier $\lambda > 0$ is:

$$L \doteq \frac{M}{\log(2)} \sum_{\forall i \in \mathcal{S}} \log \big(1 + \frac{|h(i)|^2 \cdot P(i)}{2\sigma^2}\big) + \lambda\big(E_0 - M \sum_{\forall i' \in \mathcal{S}} P(i')\big) . \tag{6}$$

The derivative[2] for the unknown transmit power P_j in block $j \in \mathcal{S}$ is:

$$\frac{\partial L}{\partial P(j)} = \frac{M}{\log(2)} \frac{1}{1 + \frac{|h(j)|^2 \cdot P(j)}{2\sigma^2}} \cdot \frac{|h(j)|^2}{2\sigma^2} - \lambda \cdot M . \tag{7}$$

Setting this derivative to zero, canceling $M \neq 0$, and setting $\tilde{\lambda} \doteq \lambda \log(2)$ produces

$$P^*(j) = \Big(\frac{1}{\tilde{\lambda}} - \frac{2\sigma^2}{|h(j)|^2}\Big)^+ . \tag{8}$$

which is a "waterfilling" solution (e.g. [4]). Equation (8) introduced a "max"-operation according to $(x)^+ \doteq \max(0, x)$ to ensure that the solutions for the powers do not take negative values; the Karush-Kuhn-Tucker [5] conditions guarantee that (8) is still an optimal solution to our problem. The value of $\tilde{\lambda}$ must be chosen such that (5) is fulfilled. The maximum number of information bits N^* is given by substituting (8) into (3) with $P(i) = P^*(i)$. For a set $\{|h(i)|^2\}$ of channel-power coefficients numerical determination of $\tilde{\lambda}$ tries a value for it, solves (4) and checks if E is close to the pre-specified value of E_0. If $E > E_0$ the algorithm increases $\tilde{\lambda}$ to get closer to E_0: iteratively repetition of this process determines the optimal solution for $\tilde{\lambda}$ for this particular sequence of channel coefficients.

[1]This paper uses the notion of an energy definition for discrete-time signals as a sum of transmit powers in the blocks, similar to a physically meaningful definition of "energy" in the analogue world. To simplify notation, however, any "time-scaling" with a transmit symbol period is omitted.

[2]To avoid confusion with the sum index i, the index is changed to j for the power for which the derivative is taken.

IV. The Fixed-Point Algorithm

The previously presented constrained search for $\tilde{\lambda}$ is not practical, since the step-size which defines by how much λ is increased or decreased in every iteration depends on the channel coefficients previously measured. This section introduces an alternative representation of the problem at hand: The problem and its solution can be rewritten equivalently by dividing (4) by M and $2\sigma^2$:

$$\frac{E/M}{2\sigma^2} = \sum_{\forall i \in \mathcal{S}} \frac{P(i)}{2\sigma^2} \leq \frac{E_0/M}{2\sigma^2} . \tag{9}$$

With the definition

$$\mathcal{S}^* \doteq \{i : |h(i)|^2 > \lambda\} , \tag{10}$$

combination of (8) and (9) leads to

$$\sum_{\forall i \in \mathcal{S}^*} \Big(\frac{1}{\lambda} - \frac{1}{|h(i)|^2}\Big)^+ \leq \frac{E_0/M}{2\sigma^2} . \tag{11}$$

Please note that now $\lambda = 2\sigma^2 \cdot \tilde{\lambda}$. With definition (10), the following expression is equivalent when requesting equality ($\overset{!}{=}$) for the energy constraint

$$\frac{E_0/M}{2\sigma^2} \overset{!}{=} \sum_{\forall i \in \mathcal{S}^*} \Big(\frac{1}{\lambda} - \frac{1}{|h(i)|^2}\Big) = \frac{|\mathcal{S}^*|}{\lambda} - \sum_{\forall i \in \mathcal{S}^*} \frac{1}{|h(i)|^2} \tag{12}$$

Solving (12) for λ reads

$$\lambda = \frac{|\mathcal{S}^*|}{\frac{E_0/M}{2\sigma^2} + \sum_{\forall i \in \mathcal{S}^*} \frac{1}{|h(i)|^2}} . \tag{13}$$

Note that (13) is only an implicit characterization of λ as the right-hand side also depends on λ (via (10)). The result for λ does, however, *not* depend on the *order* with which the summation in the denominator of (13) is calculated. Hence, in order to compute λ, the set $\mathcal{H} \doteq \{h(i), i \in \mathcal{S}\}$ of non-zero channel coefficients is sorted in descending order such that an ordered set $\hat{\mathcal{H}} \doteq \{\hat{h}(j), j = 1, 2, ..., |\mathcal{S}|\}$ with $\hat{h}(j) \geq \hat{h}(j')$ if $j < j'$ is obtained. According to (10), the set

$$\hat{\mathcal{S}}^* \doteq \{j : |\hat{h}(j)|^2 > \lambda\} . \tag{14}$$

of all index-numbers j (of the sorted channel coefficients) for which the inequality $|\hat{h}(j)|^2 > \lambda$ is fulfilled, is introduced. The *largest* of those indices is denoted by $J(\lambda)$, i.e.,

$$J(\lambda) = \max_{j : |\hat{h}(j)|^2 > \lambda} j , \tag{15}$$

so that (13) can be rewritten according to

$$\lambda = \frac{J(\lambda)}{\frac{E_0/M}{2\sigma^2} + \sum_{j=1}^{J(\lambda)} \frac{1}{|\hat{h}(j)|^2}} . \tag{16}$$

This fixed-point function is subsequently called $\varphi(\lambda)$. In the fixed-point, $\varphi(\lambda) = \lambda$. Obviously, (16) calls for the use of fixed-point iterations to find the optimal value for λ. A sensible starting point for the fixed-point iterations can be found as follows: $J(\lambda)$ will never be zero, as, even for the worst channel, the transmit energy budget will be used for at least one (the largest) channel coefficient even if almost no information can be transmitted. Based on this consideration the following procedure to compute λ is proposed:

1) **The algorithm has not yet run on a similar set of channel-power coefficients.** A start-value for $J(\lambda)$ needs to be defined: $J(\lambda) \in \{1, 2, ..., |\mathcal{S}|\}$, and simulations have shown that of the $|\mathcal{S}|$ channel-power coefficients in most cases just a few are really used for data transmission. Therefore the start-value is set to be $J(\lambda) = 1$. This avoids defining a start-value for λ as it follows from the initial choice of $J(\lambda)$ by use of (16).
2) As indicated in the previous step, with initial $J(\lambda)$ given, λ is computed by (16).
3) **If an optimum λ is known from a recent set of channel-power coefficients, the algorithm starts here and omits steps 1) and 2).** Instead, previously optimum λ is re-used to dramatically cut down the necessary number of iterations to find the new fixed-point, and therefore, the new optimum waterfilling-parameter λ, c.f. section V.
4) The new numerator in (16), $J(\lambda)$, is generated by counting the number of channel-power coefficients which satisfy $|h(i)|^2 > \lambda$.
5) The new denominator is calculated by summing all channel-power coefficients $|h(i)|^2$ with $1 \leq i \leq J(\lambda)$, and adding the constant "SNR"-term $\frac{E_0/M}{2\sigma^2}$. The algorithm carries out the division.
6) If the new value for λ and the old one are the same or within a certain range, the iterations terminate. The most recent value of λ needs to be divided by $2\sigma^2$ according to (8). The resultant $\tilde{\lambda}$ is the solution. Otherwise the algorithm continues with Step 4).

Exploration of the convergence behavior of the above fixed-point function (16) for very general assumptions concerning the statistical properties of the channel-power coefficients $|h(i)|^2$ results in a proof that can be split into the proof of two statements, the combination of which will prove the convergence of the fixed-point iteration. Due to spatial limitations (the proof takes roughly 10 pages, without many explanations or any illustrations), only the main ideas are pointed out, omitting technical derivations:

1) The proof extends the fixed-point function $\varphi(\lambda)$ to a continuously differentiable function $\tilde{\varphi}(\lambda)$ which generalizes $\varphi(\lambda)$, but has the same global characteristics (most notably the same fixed-point), adding differentiability. This smoothing is achieved by generalizing both the functions $|\hat{h}(j)|^2$ and $J(\lambda)$ to real-valued domain and real-valued codomain, respectively:

$$|\tilde{h}(j)|^2 = F^{-1}_{|H|^2}\left(1 - \frac{j}{N}\right) \qquad (17)$$

$$\tilde{J}(\lambda) = N \cdot \left(1 - F_{|H|^2}(\lambda)\right) \qquad (18)$$

Please note that now $F_{|H|^2}(\lambda)$ is a bijective function, i.e., that $f_{|H|^2}(\lambda) \neq 0, \quad \lambda \in]x, y[$, for $x < y$. Usually, $x = 0$ and $y = 1$ seems sensible for a realistic wireless scenario, not restricting generality. The smoothed function $\tilde{\varphi}(\lambda)$ coincides with $\varphi(\lambda)$ for all points where $N \cdot \left(1 - F_{|H|^2}(\lambda)\right) \in \mathbb{N}$, with N denoting the number of channel-power samples:

$$\tilde{\varphi}(\lambda) = \frac{N \cdot \left(1 - F_{|H|^2}(\lambda)\right)}{\frac{E_0/M}{2\sigma^2} + \int_0^{N \cdot \left(1 - F_{|H|^2}(\lambda)\right)} \frac{1}{F^{-1}_{|H|^2}\left(1 - \frac{j}{N}\right)} \, dj} . \qquad (19)$$

2) Exploring the behavior of $\tilde{\varphi}(\lambda)$ at $\lambda = 0$ and $\lambda = 1$ leads to to the conclusion that $\tilde{\varphi}(0) > 0$, $\frac{d}{d\lambda}\tilde{\varphi}(0) \geq 0$, $\tilde{\varphi}(1) = 0$, and $\frac{d}{d\lambda}\tilde{\varphi}(1) = 0$. Therefore, $\tilde{\varphi}(\lambda)$ has *at least one* fixed-point, since there must be at least one intersection with $f(\lambda) = \lambda$.
3) By calculating the first derivative of $\tilde{\varphi}(\lambda)$, it can be shown that every λ_i, for which $\frac{d}{d\lambda}\tilde{\varphi}(\lambda_i) = 0$ is a fixed-point.
4) It is shown that every fixed-point λ_i lies in a local maximum of $\tilde{\varphi}(\lambda)$.
5) Based on the previous findings, it is shown that the converse of 4) holds: every fixed-point λ_i has the property $\frac{d}{d\lambda}\tilde{\varphi}(\lambda_i) = 0$.
6) Concluding, based on 6), the function $\tilde{\varphi}(\lambda)$ has *at most one* fixed-point. Therefore, together with 3), it was shown that $\tilde{\varphi}(\lambda)$ has exactly one fixed-point, which is also the maximum of $\tilde{\varphi}(\lambda)$.

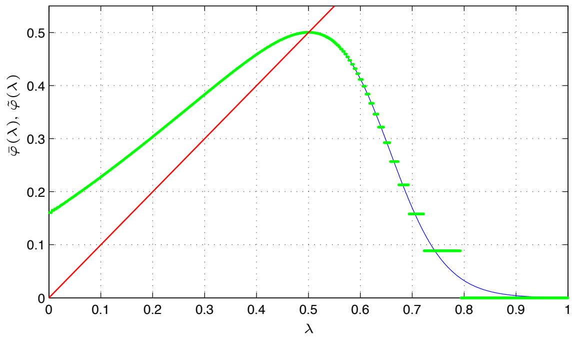

Fig. 1. Comparison of $\bar{\varphi}(\lambda)$ and $\tilde{\varphi}(\lambda)$ for a Rayleigh fading channel and Rayleigh-parameter $\sigma = 0.2$.

To see the usefulness of this approach for proving the existence and properties of a unique fixed-point, see fig. 1, illustrating analytically calculated $\tilde{\varphi}(\lambda)$ for a Rayleigh-fading channel with a Rayleigh-parameter of 0.2 and 1000 channel-power coefficients. In the figure, you see the comparison to $\bar{\varphi}(\lambda)$, numerically computed by averaging 1000 channel-power coefficients over 1000 nearest samples each. It can be observed that the two functions coincide, except for large values of λ, where $J(\lambda)$ jumps more and more due to the increasingly rare occurrence of channel-power coefficients for large λ. The assumptions on the statistical properties of the channel-power coefficients for extreme values of λ are:

$\lambda = 0$	$\lambda = 1$
$F_{\|H\|^2}(0) = 0$	$F_{\|H\|^2}(1) = 1$
$f_{\|H\|^2}(0) = 0$	$f_{\|H\|^2}(1) = 0$
$f'_{\|H\|^2}(0) \geq 0$	

Concerning $f_{|H|^2}(0) = 0$, it is assumed that only channel-power coefficients are considered which would allow for the transmission of information. For $|h|^2 = 1$, $f_{|H|^2}(1) = 0$, since it is assumed that *no* channel will conduct signals *absolutely unattenuated*.

V. Algorithmic Performance

The following code is a Matlab-Implementation providing the following basic functionality:

- Generation of 1000 Rayleigh-distributed channel-power coefficients h2 with the same statistical properties as used for all figures in this paper.
- sorting of channel-power coefficients according to size, and
- handing them over to a recursive implementation of the actual fixed-point function, which furthermore includes a counter which is incremented in each iteration of the fixed-point algorithm.

To keep the size of the code small, only the most basic functionality has been implemented - nevertheless, copying the whole code to a Matlab-File will execute as-is:

```
function Efficient_Waterfilling
  close all; clear all; clc

  SAMPLES = 1000;
  Rayleigh_Parameter = 0.2; SNR = 10;
  lambda_old = 0.3; it = 0;

  h2 = raylrnd(Rayleigh_Parameter,SAMPLES,1);
  h2_sorted = sort(h2,'descend');
  [lambda,iterations] = FP(h2_sorted,SNR,lambda_old,it)
end

function [lambda,it] = FP(h2_sorted,SNR,lambda_old,it)
  it = it+1;
  J = length(find((h2_sorted - lambda_old)>0));
  lambda = J/(SNR + sum(1./h2_sorted(1:J)));
     if abs(lambda-lambda_old)>0.00001
         [lambda,it] = FP(h2_sorted,SNR,lambda,it);
     end
end
```

Please note that the only two outputs of the code are the calculated (with an accuracy of 0.00001) waterfilling-parameter λ, and the number of iterations it took the fixed-point algorithm to converge with the given accuracy.

Of course, the number of iterations depends on the initial λ with which the fixed-point iterations are started. Fig. 2 contains the convergence traces for 1000 realizations of initial $\lambda \sim \mathcal{U}[0,1]$, where the channel-power coefficients were generated as in the Matlab-code above. It is important to note that one iteration can be avoided if the initial choice of λ does not take on values of λ for which $\varphi(\lambda) = 0$, which occurs for large values only, see fig. 1. These convergence traces are plotted in a different color in fig. 2. The values of λ, for which the first iteration is wasted do not coincide exactly with the simulation from fig. 1 , since fig. 2 is based upon one single realization rather than an average over many realizations of channel-power coefficients and is therefore just a typical, but nevertheless random realization.

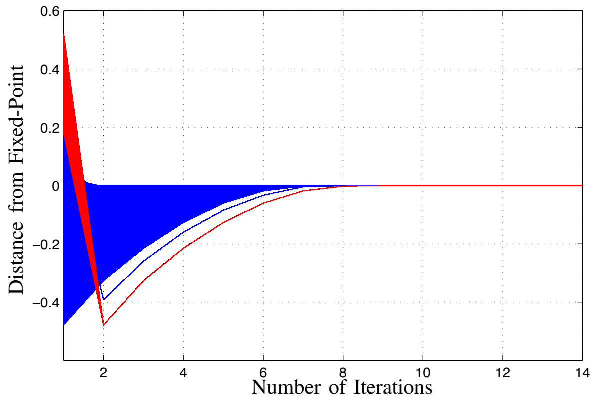

Fig. 2. Illustration of convergence behavior of fixed-point algorithm.

VI. Conclusions and Outlook

This paper has introduced a highly efficient fixed-point algorithm for calculating the optimum waterfilling-parameter which maximizes the number of transmitted bits for a given set of channel-power coefficients and a sum-energy constraint. It has demonstrated its performance, and has provided a working Matlab-implementation as a demonstration of its simplicity. The presented algorithm provides the possibility to decide on the optimum power which shall be spent in the next transmission time-slot, based on the knowledge of *previously measured* channel-power coefficients. Instead of calculating the waterfilling parameter λ for the current time-slot, taking into account 1000 *previous* channel-power coefficients, and applying this measure to determine the optimum current transmit power will lead to good results, while moving precious computation-time to less critical periods. Detailed results into this direction are however subject to current research.

References

[1] P. Palomar and J. R. Fonollosa, "Practical algorithms for a family of waterfilling solutions," *IEEE Transactions on Signal Processing*, vol. 53, no. 2, pp. 686–695, Feb. 2005.

[2] T. Cover and J. Thomas, *Elements of Information Theory*, John Wiley & Sons, Inc., 1991.

[3] N. Görtz and J. Gonter, "Information transmission over a finite-lifetime channel under energy-constraints," in *Proceedings European Wireless 2011*, Vienna, April 2011.

[4] D. Tse and P. Viswanath, *Fundamentals of Wireless Communications*, Cambridge University Press, 2005.

[5] H. W. Kuhn and A. W. Tucker, "Nonlinear programming," in *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability (available online at http://projecteuclid.org)*, Berkeley, CA, USA, July/August 1950, pp. 481–492.