Optimal Service Allocation in Multi-system Scenarios with Linear Subsystem Capacity Regions

*Ingmar Blau and Gerhard Wunder**
Fraunhofer German-Sino Mobile Communications Lab, Heinrich-Hertz-Institut
Einstein-Ufer 37, D-10587 Berlin, Germany
{blau,wunder}@hhi.fhg.de

***Abstract*— This paper discusses the user assignment in wireless multi-system multi-service scenarios. The problem is to find the optimal service-mixes on all individual subsystems. The optimal service-mixes maximize the combined system user capacity subject to a fixed combined mixture of service requests. In general directly solving the optimization problem with the combined service-mix constraint is difficult. However, using duality a reformulation is presented that can be solved efficiently with standard convex optimization tools. An algorithm is presented that converges to the optimum service allocation for linear user capacity regions. However, convergence can be proofed for arbitrary convex regions.**

***Keywords*— multi-system, multi-service, cross-system, user-capacity, convex optimization**

1 Introduction

Today mobile operators offer a multitude of services on multiple radio access technologies to their customers. Modern user equipment is often capable to support several services and radio technologies. Although new air-interfaces will conquer the market soon operators will still be interested to exploit their legacy systems. In terms of coverage, maintenance and overload situations it can be beneficial for them to support users by an air-interface of their choice flexibly. These needs are reflected in the releases of 3GPP that cover not only the possibility to perform inter-system hand-overs but also allow the exchange of load parameters. This parameters are usually used to balance the consumption of resources on individual radio access technologies and aim to increase the service reliability. In general air-interfaces can be better suited to support a specific service than other ones. One reason for this can be differences in modulation and coding schemes for example. In this case load-balancing without consideration of the service-type leads to arbitrary service-mixes on all air-interfaces and is in general suboptimal for common utility functions. For operators therefore the following question arises. What are the optimal service-mixes on individual air-interfaces, that result in a requested overall mixture of services? This questioned was addressed in [1]. There principles were discussed how multiple services should be allocated in multi-system scenarios. It was shown that high gains in terms of supportable users can be achieved if service aware user assignment is performed. [1] also presented an algorithm to find a near optimum service allocation. However, for a given combined service-mix, individual subservice-mixes have to be extracted from a mapping table.

For many radio technologies and services the number of supportable users of one service decreases linearly with the number of users of other services. This was shown for voice and data services in CDMA systems in [2] [3], but is suitable for all interference limited subsystems. In [4], [5], similar results were obtained and applied to cross-system optimization. Motivated by these results this paper presents an algorithm to calculate the optimal service-mix on all available air-interfaces assuming linear subsystem capacity regions (SSCR). It can be shown however that the algorithm is not restricted to the linear case. As long as all SSCR are convex convergence to the global optimum can be proved. While the calculation of the user assignment that maximizes a weighted sum of service arrival rates can be achieved with standard optimization tools here the maximization of rates at a fixed sum service-mix is considered. This relates to the scenario when an operator wants to support services proportionally fair to the service requests. Directly solving the optimization problem subject to a fixed sum service-mix is difficult. Therefore, based on duality, a reformulation of the problem is presented. The transformed problem can be efficiently solved with standard convex optimization tools. In contrast to the algorithm presented in [1] it is not necessary to calculate a complete mapping between combined and subsystem service-mixes. For a given combined service-mix convergence to the optimal subservice-mixes is usually achieved in few iterations.

This paper is organized as follows: In Section 2 the system description, the definition of the SSCR and the optimization problem are given. Section 3 covers general properties of combined capacity regions. Section 4 shows the dual reformulation of the optimization problem. The algorithm and simulation results are presented in Section 5. The paper is concluded in Section 6.

[1]The authors are supported in part by the *Bundesministerium für Bildung und Forschung (BMBF)* under grant FK 01 BU 566

2 System Model

Notations: Bold symbols denote vectors or matrices. The indices m, n correspond to the m^{th}, n^{th} subsystem, service and are used to address elements of vectors or matrices. $[\cdot]_i$ is the i^{th} entry of a vector, $(\cdot)^T$ denotes transpose and $\boldsymbol{I}$ is the identity matrix. $\|\cdot\|_i$ is the $\mathbf{l}_i$ norm. All used variables and entries of vectors are real and positive if not stated different.

A scenario consisting of M subsystems all supporting the same set of N service classes $c_n, n \in .\{1, \ldots, N\}$ is considered. All subsystems have the same coverage and equal channel models are used. Further it is assumed that each user uses only one air-interface per service at a time and has full connectivity to all M air-interfaces. Diversity gains achievable by assigning multiple subsystems to a single user have been shown to be small at high user numbers in [1] and are therefore not considered. In the scenario there are K users all belonging to one of the N service classes with K_n members. For simplicity the QoS constraints of service classes are given by a minimum SIR β_{c_n}.

2.1 Service-mix

The service-mix $\boldsymbol{\alpha}_m$ in subsystem $m \in 1, \ldots, M$ is a N-dimensional vector. Its entries are defined as the ratio of service arrival rate r_m^n in air-interface m and service class n and the sum of rates in that air-interface

$$\alpha_m^n = [\boldsymbol{\alpha}_m]_n = r_m^n / \sum_n r_m^n. \tag{1}$$

In the same way the N-dimensional service-mix vector $\boldsymbol{\alpha}$ of the combined systems can be defined with elements:

$$[\boldsymbol{\alpha}]_n = \sum_m r_m^n / \sum_{n,m} r_m^n \tag{2}$$

2.2 Subsystem Capacity Regions

In this paper the user capacity is considered. The subsystem user capacity region (SSCR) $\boldsymbol{C}_m$ of a subsystem m is defined as the set of all service arrival rate vectors $\boldsymbol{r}_m = [r_m^1, \ldots, r_m^N]$, where a maximum average outage probability $P_{out,max}$ is not exceeded:

$$\boldsymbol{C}_m = \{\boldsymbol{r}_m : P_{out}(\boldsymbol{r}_m) \leq P_{out,max}\} \tag{3}$$

The average outage probability is defined as the probability that not enough resources are available in an air-interface, independent of the service class, to meet a minimum SIR. Since we restrict our model to linear regions $\boldsymbol{C}_m$ can be defined by an n-dim hyperplane with normal vector $\boldsymbol{n}_m$ and distance d_m

$$\boldsymbol{C}_m = \{\boldsymbol{r}_m : \boldsymbol{n}_m^T \boldsymbol{r}_m \leq d_m\}. \tag{4}$$

Equivalently $\boldsymbol{C}_m$ can be characterized by the set of N maximum single service rates $r_{m,max}^n$ which are the intersection points of the hyperplane with the axes.

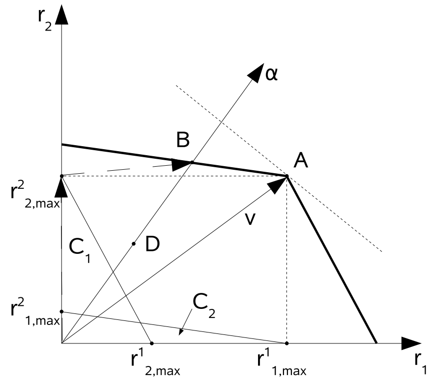

Fig. 1. Characterization of capacity region

2.3 Combined SSCR and Optimization Problem

In multi-system scenarios subsystems are usually orthogonal to each other. Therefore the combined system capacity is comprised by the sum over all SSCR $\boldsymbol{C} = \sum_m \boldsymbol{C}_m$. Now the optimization problem can be formulated. For a given combined service-mix $\boldsymbol{\alpha}$ maximize the feasible sum rate and determine the corresponding optimal service-mixes $\boldsymbol{\alpha}_m$ on all subsystems

$$\begin{aligned} C^{\alpha} = \max_{\boldsymbol{\alpha}_m} & \|\sum_m \boldsymbol{r}_m\|_2 \\ \text{s.t } \boldsymbol{n}_m^T \boldsymbol{r}_m & \leq d_m, \ \forall m \in \{1, \ldots, M\} \\ \sum_m \boldsymbol{r}_m & = c\, \boldsymbol{\alpha}, \end{aligned} \tag{5}$$

with c a positive constant. Fig. 1 characterizes the capacity regions C_1, C_2 and $\boldsymbol{C}$ for $M, N = 2$. Point B is the solution to (5) at a given combined service-mix $\boldsymbol{\alpha}$. The difficulty of the optimization problem is the constraint on the combined service-mix.Without this the problem would reduce to maximizing a sum of weighted rates and could be solved with standard convex optimization tools. If one neglects the service-mix constraint and uses $\boldsymbol{\mu} = \boldsymbol{\alpha}$ as a weight vector instead, maximizing the weighted sum of service arrival rates leads to point A. As can be seen the service-mix at point A is in general not equal to the weight vector and therefore also different from point B. Point D visualizes the maximum rate vector with equal service-mixes $\boldsymbol{\alpha}$ on all SSCR, which is clearly suboptimal.

3 Properties of the Individual and Combined Capacity Regions

3.1 Individual SSCR

In this paper linear SSCR are assumed. Since this assumption is not intuitive some remarks and simulations are presented here.

For interference limited systems like CDMA for example the limiting factors determining the outage and therefore the SSCR are usually power constraints.

Although it is noted here that system capacity is limited also with infinite power budgets and outage can also occur because of a lack of codes. For CDMA uplink a user $i \in \{1, \dots, K\}$ with minimum SIR β_i is feasible if

$$\beta_i \leq \frac{h_i p_i}{\frac{1}{N_{SF}}(1+f)\sum_{j \neq i} h_j p_j + n_{th}}, \qquad (6)$$

with spreading gain N_{SF}, channel amplification h_i, transmission power p_i and n_{th} AWGN. Intercell interference is modeled as multitude of the intracell interference and incorporated by using factor f. Assuming that all users belong to N service classes c_n with K_n users and $\beta_{i \in c_n} = \beta_{c_n}$ the equation above can be reformulated following [2] to:

$$\sum_{n=1}^{N} K_n \frac{1}{\frac{\delta}{\beta_{c_n}} + 1} \leq 1 - \frac{n_{th}\delta}{\min_{i \in I}[h_i p_i(\frac{\delta}{\beta_i} + 1)]} \qquad (7)$$
$$\delta = N_{SF}(1+f)$$

This equation describes the area below a hyperplane were the normal vector is determined by the left side of the equation and the distance by the right one. To obtain the capacity region one has to evaluate the probability that the condition above is met. Under the assumption that the right side of the equation is independent on the service-mix the capacity region for CDMA is also limited by a hyperplane with a normal vector proportional to the left side of (7).

In GSM outage results most times from a lack of free time slots. In this case the capacity region is limited also by a hyperplane with normal vector pointing in a direction proportional to the number of time-slots needed per service. However, at high data rates it also occurs because of too low SIR and therefore lack of power.

Without simplifications the form of capacity regions depends on the call arrival processes and channel distributions and is in general not linear. For low outage probabilities simulations however justify the approximations in form of hyperplanes for GSM and UMTS. The simulated capacity regions for UMTS and GSM are shown in Figure 2 for two service classes with minimum data rates of 12.2 kbit/s and 22 kbit/s respectively and a maximum outage probability of $P_{out,max} = 0.1$. The simulation environment considers Poisson service arrival processes with exponentially distributed duration with $120s$ mean, equally distributed users on 30 cells, lognormal distributed slow fading and perfect power control for UMTS. Similar results have been obtained in [4] for UMTS/GSM and, using noise rise as outage condition, in [3] for CDMA.

3.2 Combined SSCR and weighted sum optimization

Some basic properties of combined sum capacity regions with linear subregions can be formulated:

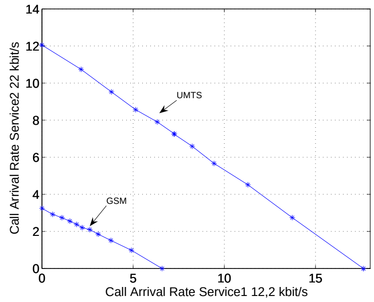

Fig. 2. Simulated SSCR for UMTS and GSM

1) The combined capacity region will be a N-dimensional convex polytope.
2) To reach any point on the boundary there exists always a set of rate vectors where for $M - N + 1$ air-interfaces the rate vector has only one nonzero entry

$$[\boldsymbol{\alpha}_m]_i = \begin{cases} 1 \ i = j \\ 0 \ i \neq j \end{cases} m \in \mathbb{M}' \subseteq [1, \dots, M]. \qquad (8)$$

The proof relies on Lagrange theory and is omitted due to the lack of space.
3) To determine the weighted sum rate (point A in Fig. 1) knowing the normal vectors $\boldsymbol{n}_m$ is sufficient.

Finding the maximum weighted sum rate with a weight vector $\boldsymbol{\mu}$ can be interpreted geometrically as maximizing the distance of a hyperplane with normal vector $\boldsymbol{\mu}$ to the origin over $\boldsymbol{C}$. The maximum weighted sum rate vector $\boldsymbol{v}$ to the osculation point A is the sum of the vectors $\boldsymbol{v}_m$. For all SSCR $\boldsymbol{v}_m$ is obtained by maximizing the distance of the hyperplane over individual subregions and taking the rate vector to the osculation points. For linear sub capacity regions the osculation points can be determined by

$$v_m^i = \begin{cases} r_{m,max}^i \text{ if } i = \arg\max_n n_m^n \mu_n \\ 0 \ else \end{cases} \qquad (9)$$

If all $\boldsymbol{v}_m$ are unique point A will be a vertex, if multiple solutions exist A is a rim or a face of the convex polytope $\boldsymbol{C}$. The maximum weighted sum of service arrival rates can be written as

$$\begin{aligned} f_{WSR}(\boldsymbol{\mu}) &= \max_{\boldsymbol{r}_m \in \boldsymbol{C}_m} \sum_{m=1}^{M} \boldsymbol{\mu}^T \cdot \boldsymbol{r}_m \\ &= \|\sum_{m=1}^{M} \boldsymbol{v_m}(\boldsymbol{\mu})\|_2, \end{aligned} \qquad (10)$$

which is shown in the following. If the SSCR are known one element of the rate vectors $\boldsymbol{r}_m$ can be expressed in terms of the other ones

$$r_m^1 = f_m(r_m^2, \dots, r_m^N). \qquad (11)$$

Using the constraints that for $n' = \{2\dots,N\}$, $0 \leq r_m^{n'} \leq r_{m,max}^{n'}$ the Lagrange function can be expressed by

$$L(r_m^{n'},\lambda) = \sum_{n=2}^{N}\sum_{m=1}^{M}[r_m^n(\mu_n + \lambda_m^n - \lambda_{M+m}^n) + \lambda_{M+m}^n r_{m,max}^n] + \mu_1 \sum_{m=1}^{M} f_m(r_m^2,\dots,r_m^N), \tag{12}$$

where $\lambda_m^n, \lambda_{M+m}^n$ are the dual variables. For convex SSCR at the optimum the KKT conditions are satisfied:

$$\frac{\partial L(r_m^{n'},\lambda)}{\partial r_m^{n'}} = \mu_n + \lambda_m^n - \lambda_{M+m}^n + \frac{\partial}{\partial r_m^{n'}}\mu_1 f_m(\dots) = 0$$
$$\forall_m^{n\in n'} \lambda_m^n r_m^n = \lambda_{M+m}^n(r_m^n - r_{m,max}) = 0; \lambda \geq 0 \tag{13}$$

For linear SSCR $\frac{\partial}{\partial r_m^{n'}} f_m(r_m^2,\dots,r_m^N)$ is constant and determines if r_m^n is zero, $r_{m,max}^n$ or can be chosen freely. For arbitrary convex SSCR the slope of the border of the capacity regions are equal at the optimum service-mix or r_m^n is zero or $r_{m,max}^n$ respectively.

4 Maximum Capacity with fixed service-mix

The maximum weighted sum rate point on the combined capacity region $\boldsymbol{C}$ and the corresponding rate vectors on the SSCR can now be calculated by evaluating equation (9) and (10) for a given weight vector. One way to solve the optimization problem in (5) is to evaluate (9) and (10) for all possible weight vectors and create a map containing all vertices of the convex polytope. Then exhaustive search can be used to determine the face of the polytope, where a vector in direction of the combined service-mix pierces. However, with increasing M, N this strategy gets very elaborate and seems to be infeasible if SSCR are arbitrary convex regions. Therefore a different approach to solve the optimization problem is presented here. The first step is to determine the normal vector $\boldsymbol{\mu}^*$ of the face of the polytope on which B lies. Geometrically one could unroll a hyperplane on $\boldsymbol{C}$ in the direction of $\boldsymbol{\alpha}$ until N neighboring vertices of B are determined and calculate $\boldsymbol{\mu}^*$. Mathematically a more elegant method using duality is presented here. Although the utility function and all constraints of (5) are convex directly solving the problem is difficult. However some reformulations simplify the problem. At first the service-mix constraint can be incorporated in the following way:

$$C^\alpha = \max_{\boldsymbol{r}_m \in \boldsymbol{C}_m} \min_n [\boldsymbol{\alpha}]_n^{-1} \sum_m r_m^n. \tag{14}$$

The minimum over all services can also be written in terms of

$$C^\alpha = \max_{\boldsymbol{r}_m \in \boldsymbol{C}_m} \min_{\sum_{n=1}^N \mu_n = 1} \sum_{n=1}^{N} \mu_n [\boldsymbol{\alpha}]_n^{-1} \sum_m r_m^n. \tag{15}$$

Since $\boldsymbol{C}_m$ are convex and compact subsets of $\mathbb{R}^N$ assumed to be non empty and the same is true for the set of weight vectors, the min-max Theorem (proposition 5.4.4 in [6]) can be applied, resulting in

$$C^\alpha = \min_{\sum_{n=1}^N \mu_n = 1} \max_{\boldsymbol{r}_m \in \boldsymbol{C}_m} \sum_{n=1}^{N} \mu_n [\boldsymbol{\alpha}]_n^{-1} \sum_m r_m^n. \tag{16}$$

In [7] a similar result was obtained by directly applying Lagrange Theory. In (16) one recovers the maximum weighted sum rate problem which was solved in the last section. The maximum weighted sum rate has then to be minimized over $\boldsymbol{\mu}$. For this problem standard convex optimization tools can be used to find the weights that minimize (16). The optimum weights determine the normal vector of the wanted face of the polytope.

$$\boldsymbol{\mu}^* = \arg\min_{\sum_{n=1}^N [\boldsymbol{\mu}]_n = 1} f_{WSR}(\boldsymbol{\gamma}), \tag{17}$$

with $[\boldsymbol{\gamma}]_i = [\boldsymbol{\mu}]_i / [\boldsymbol{\alpha}]_i$. With $\boldsymbol{\mu}^*$ and $\boldsymbol{\alpha}$ the optimum point B in Fig. 1 can easily be determined. The optimal subsystem rate vectors and therefore the service-mixes can now be calculated from (9) to $\boldsymbol{v}_m^* = \boldsymbol{v}_m(\boldsymbol{\mu}^*)$. For all non unique $\boldsymbol{v}_m(\boldsymbol{\mu}^*)$ the service-mix can be obtained by solving a system of linear equations.

5 Ellipsoid Method

To solve (16) like in [7] the ellipsoid method is used here. It is a generalization of the bisection method to multiple dimensions. Since the minimization is performed over the surface of a polytope derivatives do not need to exist. However, since $\boldsymbol{C}$ is continuous everywhere, a sub-gradient can always be found. In connection with convexity of the polytope this sub-gradients can then be used with the ellipsoid method and convergence to the optimum can be proofed. A valid sub-gradient, $\sum_m ([\boldsymbol{\alpha}]_n^{-1} r_m^n - [\boldsymbol{\alpha}]_N^{-1} r_m^N)$, $n \in \{1,\dots,N-1\}$ can be read off if the sum constraint on the weights is integrated in (16) and reformulated in the following way

$$\begin{aligned} C^\alpha &= \min_{\boldsymbol{\mu}} \max_{\boldsymbol{r}_m \in \boldsymbol{C}_m} \sum_{n=1}^{N-1} \mu_n [\boldsymbol{\alpha}]_n^{-1} \sum_m r_m^n \\ &\quad + (1 - \sum_{n=1}^{N-1} \mu_n)[\boldsymbol{\alpha}]_N^{-1} \sum_m r_m^N \\ &= \min_{\boldsymbol{\mu}} \max_{\boldsymbol{r}_m \in \boldsymbol{C}_m} \sum_{n=1}^{N-1} \sum_m ([\boldsymbol{\alpha}]_n^{-1} r_m^n - [\boldsymbol{\alpha}]_N^{-1} r_m^N)\mu_n \\ &\quad + [\boldsymbol{\alpha}]_N^{-1} \sum_m r_m^N, \end{aligned} \tag{18}$$

which can be proven to be an affine underestimate. The algorithm is initiated by building an $N-1$ dimensional ellipse with matrix $\mathbf{E} = (1 - \frac{1}{N})\mathbf{I}_{N-1}$ covering the feasible weight space $\sum_{n=1}^{N-1} \mu_n \leq 1$. In each iteration

Algorithm 1 Ellipsoid Method

initialize $\mathbf{E} = (1 - \frac{1}{N})\mathbf{I}_{N-1}, \mu_n = \frac{1}{N}$

while $\max eig(\mathbf{E}) >$ stopping threshold **do**

(1) calculate $\sum_m v_m(\boldsymbol{\mu})$ according to (9)

(2) calculate and normalize sub-gradient

$$\begin{aligned}[s]_n =& [\boldsymbol{\alpha}]_n^{-1} \sum_m v_m^n(\boldsymbol{\mu}) - [\boldsymbol{\alpha}]_N^{-1} \sum_m v_m^N \\ \tilde{s} =& \frac{\mathbf{s}}{\sqrt{\mathbf{sEs}}}\end{aligned} \tag{19}$$

(3) update ellipse

$$\begin{aligned}\mu_n^+ =& \mu_n - \frac{1}{N}\boldsymbol{E}\tilde{\boldsymbol{s}}, \ \forall n = \{1, \ldots, N-1\} \\ \mu_N^+ =& 1 - \sum_{n=1}^{N-1} \mu_n^+ \\ \mathbf{E}^+ =& \frac{(N-1)^2}{(N-1)^2 - 1}\left(\boldsymbol{E} - \frac{2}{N}\boldsymbol{E}\tilde{\boldsymbol{s}}\tilde{\boldsymbol{s}}^T\boldsymbol{E}\right)\end{aligned} \tag{20}$$

end while

(16) is evaluated with a weight vector corresponding to the center of the ellipse. With the resulting rate vector the sub-gradient is calculated and one half-space can be erased from the set of the feasible weight vectors. Then the ellipse with the smallest volume, which covers the intersection of the old one and feasible halfspace, is used for the next iteration step. In each iteration the volume of the ellipsoid decreases. For the algorithm there are several stopping criteria known. In the simulations the algorithm is terminated if the longest half-axis of the ellipsoid is smaller than a certain threshold. Details are given in Algorithm 1. A convergence proof and more sophisticated stopping criteria can be found in [8].

Figure 3 shows the optimization result for $N = 3$ and 30 subsystems with random SSCR. Here also the optimal rate vectors and $\boldsymbol{\alpha}$ are depicted. The convergence speed of the algorithm is shown in Fig. 4 for the weight vector elements and the maximum half-axis of the ellipse.

6 Conclusion

We presented an efficient algorithm to find the optimal service-mixes in multi-system multi-service scenarios. The aim was to maximize the combined user capacity at a given combined service-mix for linear SSCR. The min-max theorem could be applied to reformulate the optimization problem into minimizing the maximum weighted sum rate. After presenting the solution to the maximum weighted sum rate problem the ellipsoid method was applied to solve the original problem. The results were validated by simulations.

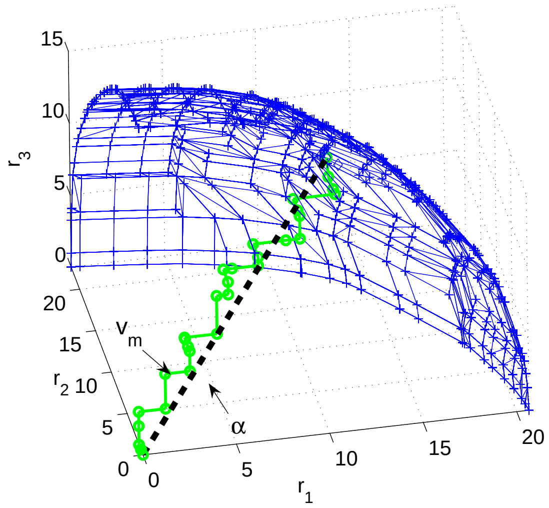

Fig. 3. Combined Capacity Region for 3 services and 30 sub-systems

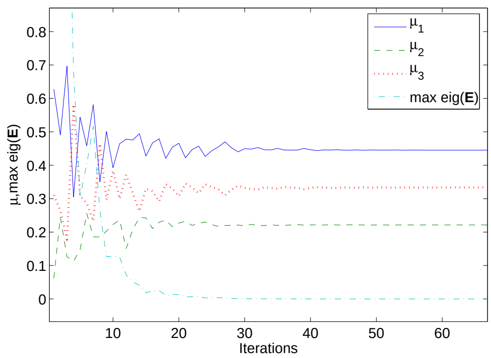

Fig. 4. Convergence speed

References

[1] Anders Furuskär and Jens Zander, "Multiservice Allocation for Multiaccess Wireless Systems," *IEEE Transactions on Wireless Communications*, vol. 4, no. 1, January 2005.

[2] Ashwim Sampath, P. Sarath Kumar, and Jack M. Holtzman, "Power Control and Resource Management for a Multimedia CDMA Wireless System," in *Personal, Indoor and Mobile Radio Communications IEEE*, 1995, vol. 1.

[3] Ashwin Sampath, Narayan B.Mandayam, and Jack M. Holtzman, "Erlang Capacity of a Power Controlled Integrated Voice and Data CDMA System," in *Vehicular Technology Conference, 1997 IEEE 47th*, 1997, vol. 3.

[4] Anders Furuskär, *Radio Resource Sharing and Bearer Service Allocation for Multi-Bearer Service, Multi-Access Wireless Networks*, Ph.D. thesis, Royal Institute of Technology, Sweden, 2003.

[5] Gabor Fodor, Anders Furuskär, and Johan Lunsjö, "On Access Selection Techniques in Always Best Connected Networks," in *ITC Specialist Seminar on Performance Evaluation of Wireless and Mobile Systems*, University of Antwerp, Antwerp, Belgium, 2004.

[6] Dimitri P. Bertsekas, *Nonlinear Programming*, Athena Scientific, Belmont, Massachusetts, 2rd. edition, 1995.

[7] Juyul Lee and Nihar Jindal, "Symmetric Capacity of MIMO Downlink Channels," in *IEEE International Symposium on Information Theory*, 2006.

[8] Stephen Boyd, *Lecture Notes on EE364b Convex Optimization II*, Stanford University, 2006.