A Robust, Tractable Analytical Foundation for Radio Resource Management: *Centralized and Decentralized Optimizations involving Data and Media Communications*

Virgilio Rodriguez

Some discussed work co-authored with D.J. Goodman and Y. Wang

ECE Department
Polytechnic University
Brooklyn, NY
email: vr@ieee.org
http://pages.poly.edu/~vrodri01/research

Outline

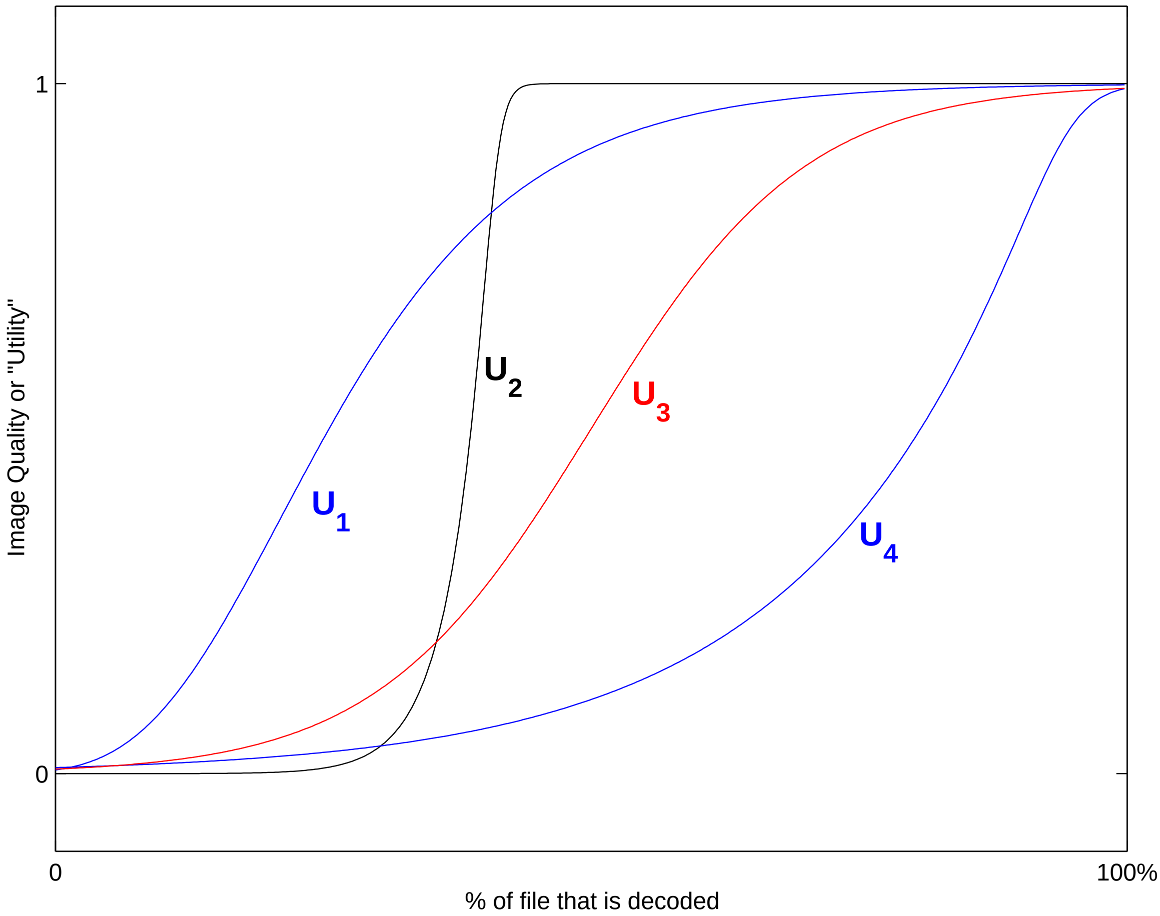
- Analytical Core
  - A tractable abstraction of the physical layer
  - A tractable abstraction of the human visual system
  - A fundamental result
- Single-user applications: data, image, video
- Decentralized multi-user applications:
  - Game formulation
  - Mechanism design
- Centralized data throughput maximization
  - Without noise
  - With noise and media terminals present

CORE: Motivation

- Many radio-resource optimizations share a common analytical core
- This core enables robust and tractable analysis
- and provides clear answers in fairly general scenarios
- Problems to which this framework applies:
  - Decentralized power control for 3G CDMA
  - Data rate and power allocation for maximal cell throughput when data and media terminals share a CDMA cell
  - Power and coding rate choice for scalable media files (images, video)
  - Choosing the “right amount” of distortion when less distortion means a higher cost
- Typically, to solve these problems one has to know how to maximize $f(x)/x$ with f an “S-curve”.

CORE: S-curves: why??

- S-curves play an essential role in the core
- For any physical layer, the function giving the probability that a data packet is received successfully as function of the SIR is an S-curve
- An S-curve can also model the perceptual quality of a reproduced image or video as function of the coding rate
- An arbitrary S-curve includes as special cases
  - an arbitrary convex curve
  - an arbitrary concave curve
  - an arbitrary threshold (step)
  - a straight line (almost)


1
0
Image Quality or "Utility"
U1
U2
U3
U4
0
100%
% of file that is decoded

CORE: An abstraction of the physical layer

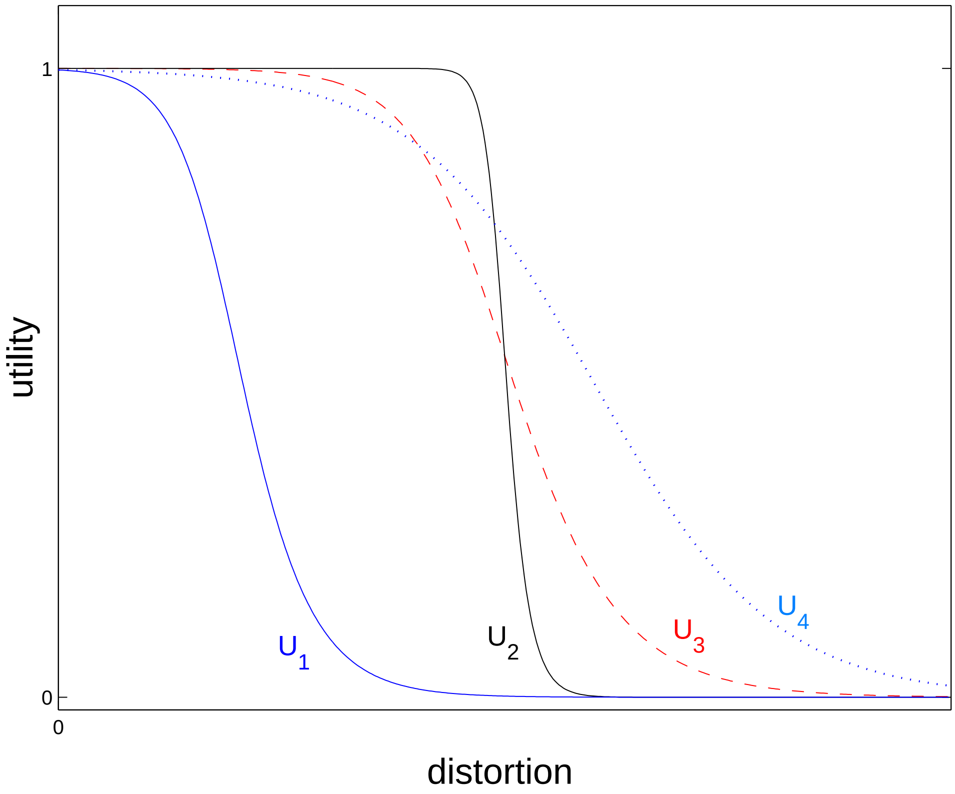
- For resource-management purposes, a frame success function (FSF) encapsulates the essential information about the physical layer
- The FSF gives probability that a data packet is received successfully as function of SIR at receiver
- It's determined by the details of the physical layer: modulation, diversity, FEC, etc.
- Ex: for Gaussian channel, non-coherent FSK modulation, with packet size M=80, no FEC, independent bit errors, and perfect error detection, the FSF is $f_s(x) = \left[1 - \frac{1}{2}\exp\left(-\frac{x}{2}\right)\right]^{80}$
- If the analysis assumes that all that is known about the FSF is that it is a smooth "S-curve", the analysis de facto accommodates most physical layers of interest

CORE: Abstracting the human visual system

- Modern image/video encoders (e.g., JPEG 2000, MPEG-4, SPIHT) produce "scalable" files (can be truncated and decoded)
- decoding complete file $\Rightarrow$ maximal quality
- decoding fewer bits per file $\Rightarrow$ lower quality
- For y the # of bits in *truncated* file, need $U(y)$ as the end-user "quality" or "utility" of decoded media.
- U must be increasing, but with which "shape" (convex, concave, linear, etc.) ?
- U can be obtained by psychophysical experiments for a specific user.
- If the analysis assumes that, all that is known about U is that it is a smooth S-curve, the analysis de facto accommodates a convex (U_4), concave (U_1), or "step" (U_2) curve! (and even a straight line!)

CORE: An abstraction of the HVS II

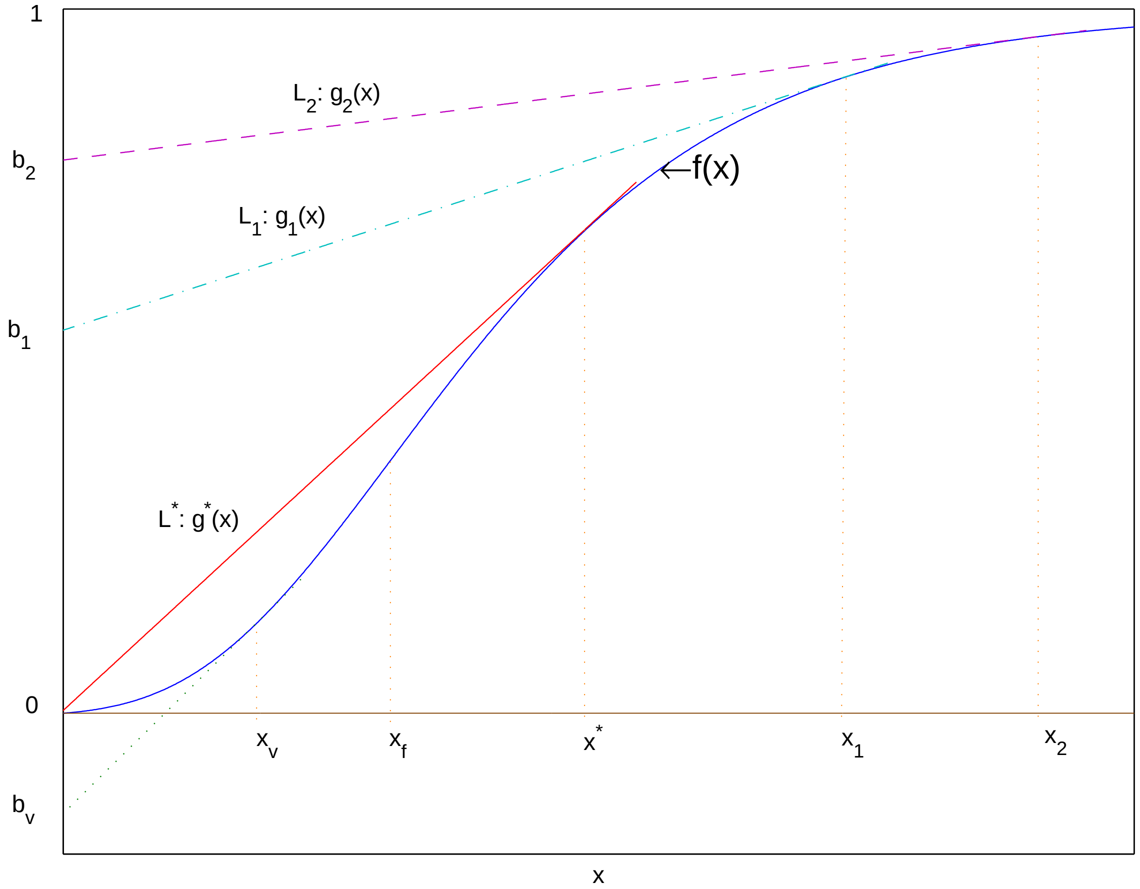
- A media signal can be useful to an end-user under various degrees of noticeable distortion
- A utility function of distortion captures this mathematically and enables the analysis of interesting trade-offs (e.g. quality vs. quantity)
- As function of distortion, U must be decreasing,..., but with which "shape" (convex, concave, linear, etc.) ?
- The literature typically assumes that, up to a level, distortion has no effect on signal quality, but beyond that level it makes the signal useless ("hard threshold").
- By assuming that the utility function is a "reversed" S-curve, the "hard threshold" is contained as a special case. Additionally, other possibilities are considered ("almost" convex, "almost" concave, etc.)
- For further details on this approach, and how it can be applied to an interesting problem see additional slides and/or a complete paper.



Beauty is in the eyes of the beholder

CORE: Maximizing S(x)/x

- Maximize $f(x)/x$ where *all that is known* about f is that its graph is an S-curve.
- No functional form ("equation") is imposed
- Sigmoidness $\Rightarrow$ f "starts out" convex at the origin, and "smoothly" transitions to concave as it approaches a horizontal asymptote
- Maximizer must solve $xf'(x) = f(x)$.
  Solution:
  - always exists
  - is unique
  - can be graphically described by drawing a tangent
- Ratio $f(x)/x$ is quasi-concave (enables application of Debreu's and other results)


1
L2: g2(x)
b2
←f(x)
L1: g1(x)
b1
L*: g*(x)
0
xv
xf
x*
x1
x2
bv
x

CORE:Recap

- Many radio-resource problems involve maximizing $f(x)/x$, with f certain monotonic function
- If all that is known about f is that it is an S-curve f can be (i) “mostly” concave, (ii) “mostly” convex, (iii) “mostly” straight, (iv) a “step”
- An S-curve can yield a useful abstraction of (i) the physical layer of a wireless communication system and (ii) the human visual system
- With f an S-curve, the solution to maximizing $f(x)/x$ is unique and can be clearly described by drawing a tangent.
- Several interesting problems involving coding rate, data rate and power allocation can be solved this way.
- For analytical work, specifying an “equation” for f typically is neither necessary nor helpful. For numerical illustrations and simulations, there are well known parametric equations yielding S-curves (e.g., Richard’s curve)

Max. bits per Joule for data transmission

- Many delay-insensitive packets to be transferred.
- Limited energy budget.
- How much power to use? Too little power $\longrightarrow$too many re-transmissions; too much power $\longrightarrow$wasted energy (e.g. increasing power by "a lot" may only increase probability of packet success by "a little")
- Solution: max. total number of bits successfully transmitted with available energy; i.e.,: maximize bits per Joule
- Analysis leads to maximizing: $R\left(f_s(x) - f_s(0)\right)/P$

$$f_s \longrightarrow \text{FSF}, \quad x \longrightarrow : \text{SIR}, \quad R \longrightarrow \text{bit rate}, \quad P \longrightarrow \text{power}$$

- It's proportional to $(f_s(x) - f_s(0))/x \equiv f(x)/x$
- The solution can be obtained by drawing a tangent from origin.

Scalable Image Files (SIF) on wireless link

- Many image files to transfer over wireless link
- Each file is "scalable" (can be truncated and decoded; e.g., JPEG 2000)
- Energy is limited!
- Assumes S-curve $u(y)$ gives quality arising from y-bit truncated file
- Transferring complete file $\Rightarrow$ maximal quality, BUT few transferred images. Transferring few bits per file $\Rightarrow$ many images transferred BUT each of low quality
- Problem: how many bits per file to transfer (where to truncate) AND at which power to transmit?
- What to do??
  Maximize total utility: $n \times u(y)$ with $n = E/c(y)$, E the available energy and $c(y)$ the energy cost of successfully transferring a y-long file

SIF: Solution

- Analysis $\Rightarrow$ with x the SIR, and R the bit rate :

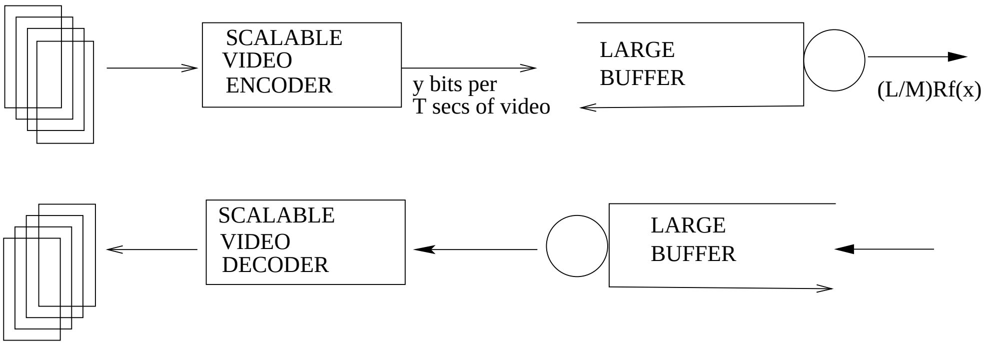
$$\max \underbrace{kR\frac{f(x)}{x}}_{\text{bits/Joule}} \times \underbrace{\frac{U(y)}{y}}_{\text{quality/bit}} \Longrightarrow \max \frac{\text{Quality}}{\text{Joule}}$$

- Analysis $\Rightarrow$ $f(x)/x$ and $U(y)/y$ can be maximized independently
- Both f and U are S-curves
- The desired solution can be obtained by drawing a tangent from origin to S-curve.
- To consider multiple users, set up game in which each terminal maximizes quality per Joule

Wireless Scalable video streaming (SVS)

- Each T-secs of video yields "scalable" file (can be truncated and de-coded; e.g., MPEG-4, SPIHT-3D)
- Assumes S-curve $u(y)$ gives segment quality from y-bit truncated file
- Energy E is limited!
- File for given segment must be transferred in a deadline of Δ secs.
- Transferring each file complete $\Rightarrow$ maximal quality per segment BUT short total viewing time with available energy. Transferring few bits per file $\Rightarrow$ long running time BUT low quality per segment.
- Problem: how many bits per file to transfer (where to truncate) AND at which power to transmit?
- What to do?? Maximize total utility: $n \times u(y)$ with $n = E/c(y)$ with $c(y)$ the energy cost of successfully transmitting a y-long file in Δ secs.

SVS: System for Wireless Scalable Video

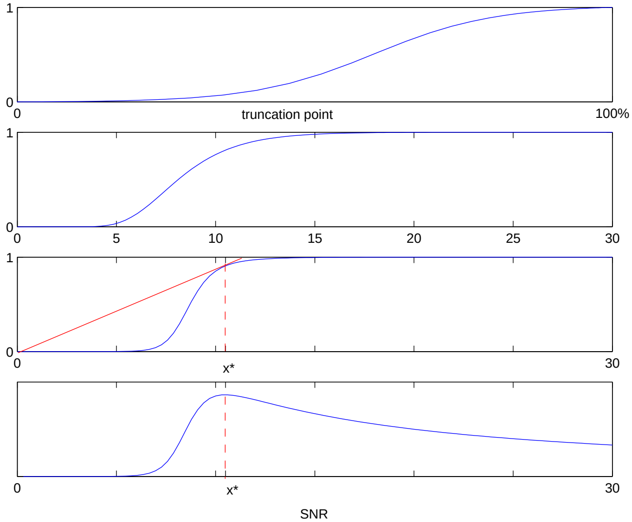


Schematic of the wireless transmission of scalably encoded live video. As first approximation, assume channel is "pseudo-deterministic" delivering $(L/M)Rf(x)$ correct information bits per sec. R is the raw bit rate, $f(x)$ the frame-success rate, and L/M the ratio of information bits to the packet size.

SVS: Solution

$$
\begin{array}{lcl}
\max\limits_{x,y} \dfrac{u(y)}{x} & & \max\limits_{x} \dfrac{u(Bf(x))}{x} \\
\text{s.t. } y = Bf(x) & \text{OR} & \text{s.t. } 0 \le x \le \bar{x} \\
0 \le x \le \bar{x} & &
\end{array}
$$

$B = (L/M)R\Delta$ interpreted as the maximum amount of information bits ("best case scenario") that can be transferred in the deadline Δ.



From the top, (i) the S-curve $u(y)$ giving the perceptual quality of a video segment, as a function of the coding rate, (ii) $f(x)$, the probability of successful reception of a packet as a function of the SIR, (iii) the composite function $u(Bf(x)) := s(x)$, (iv) the ratio $s(x)/x$ which the terminal should maximize.

Decentralized multiuser analysis (DMA): What is a Game?

- A game: each of several players chooses a “strategy” in order to receive a “payoff”.
- Payoffs depend on the choices of ALL players
- Each player is “selfish”
- Key solution concept: Nash equilibrium.
  An allocation (a strategy per player) such that no player would gain by unilaterally changing strategy (“deviating”)
- Nash equilibria are generally “inefficient”

DMA: Power Control Game

- Players: CDMA data transmitting terminals.
- Strategy: transmission power level
- Payoff : number of bits successfully transmitted per unit energy (bits/Joule)
- Signal-to-interference ratio determines bits/Joule
- A Nash equilibrium generally exists
- Equilibrium power levels are “too high”
- Challenge: how to get selfish terminals to choose lower power levels “on their own”
- For further details on this game see this WCNC-03 paper .

DMA: Mechanism design for decentralized efficiency

- "mechanism" : a set of procedures, penalties and rewards designed to guide selfish entities toward a desired outcome
- Example of a simple and useful mechanism:
  Vickery's Second Price Auction
  - Each player chooses an amount of money to bid for an object and highest bidder gets object
  - But highest bidder pays second-highest bid
  - Each player's best response is to bid its true valuation of object: "truth-telling" is optimal

DMA: The Compensation Mechanism

- Proposed by Varian in a general context
- Requires a "transferable good", say money, with which agents compensate each other.
- Assuming only 2 terminals, and that terminal 1 interferes with terminal 2 but *not* vice-versa (SIC decoding), it works as follows
  - Terminal 2 declares the amount money (or transferable good) it wishes to charge terminal 1 as compensation for each unit of interference.
  - Terminal 1 (interferer) declares the price it *offers* to pay terminal 2 as compensation.
  - The interferer (#1) must pay penalty if its offered price is different from terminal 2's price

DMA: Why does the mechanism work?

- To avoid the penalty, generally the interferer will offer to pay the exact amount terminal 2 wants.
- But why doesn't terminal 2 ask "too much"?
  - If price paid to terminal 2 exceeds its "true cost", then it "makes a profit" per unit of interference.
  - But then, it would want more interference!
  - To get the interferer to produce more, terminal 2 must lower its price.
  - Thus, at equilibrium, terminal 2 price equals its true cost, which is the "fair thing" to do.
- The mechanism also works when both terminals interfere each other, and with many mutually interfering terminals.
- For further details see additional slides and/or an extended abstract.

Centralized Throughput Max. (CTM)

- CDMA Single Cell Data Comm.
- N terminals send data to a base station
- R_c: chip rate ; R_i : data rate ; $G_i = R_c/R_i$: Spread Gain
- $f_s(\gamma_i)$: probability of correct reception of a data packet, in terms of received SIR .
- $\gamma_i := G_i \alpha_i$ is the SIR with α_i the CIR given by

$$\alpha_i = \frac{h_i P_i}{\sum_{\substack{j=1 \\ j \neq i}}^{N} h_j P_j + \sigma^2} := \frac{Q_i}{\sum_{\substack{j=1 \\ j \neq i}}^{N} Q_j + \sigma^2}$$

- h_i: gain factor (“path loss”)
- $h_i P_i := Q_i$: received power

CTM: Optimization Model

Maximize

$$\frac{f(G_1\alpha_1)}{G_1} + \frac{\beta f(G_2\alpha_2)}{G_2}$$

subject to

$$\begin{aligned} \alpha_1\alpha_2 &= 1 \\ G_1 &\geq G_0 \\ G_2 &\geq G_0 \end{aligned}$$

CTM: First-Order Conditions (FONOC)

- Augmented objective function: $\phi(G_1, G_2, \alpha_1, \alpha_2) =$

$$\frac{f(G_1\alpha_1)}{G_1} + \frac{\beta f(G_2\alpha_2)}{G_2} + \lambda(\alpha_1\alpha_2 - 1) + \sum_{i=1}^{2} \mu_i(G_0 - G_i)$$

- FONOC ($\gamma_i = G_i\alpha_i$)

$$\begin{bmatrix} (\gamma_1 f'(\gamma_1) - f(\gamma_1))/G_1^2 - \mu_1 \\ \beta(\gamma_2 f'(\gamma_2) - f(\gamma_2))/G_2^2 - \mu_2 \\ f'(\gamma_1) + \lambda\alpha_2 \\ \beta f'(\gamma_2) + \lambda\alpha_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\text{with} \begin{cases} \alpha_1\alpha_2 = 1 \\ \mu_1(G_0 - G_1) = 0 \\ \mu_2(G_0 - G_2) = 0 \end{cases}$$

CTM: An “interior” solution to FONOC

- Assume $\mu_1 = \mu_2 = 0$ and verify whether a feasible solution to FONOC consistent with these hypothesis actually exists. This allows $G_1 \geq G_0$ and $G_2 \geq G_0$.
- A closed-form “interior” solution is found

$$G_1\alpha_1 = G_2\alpha_2 = \gamma_0$$

$$\alpha_1 = \frac{1}{\alpha_2} = \sqrt{\beta}$$

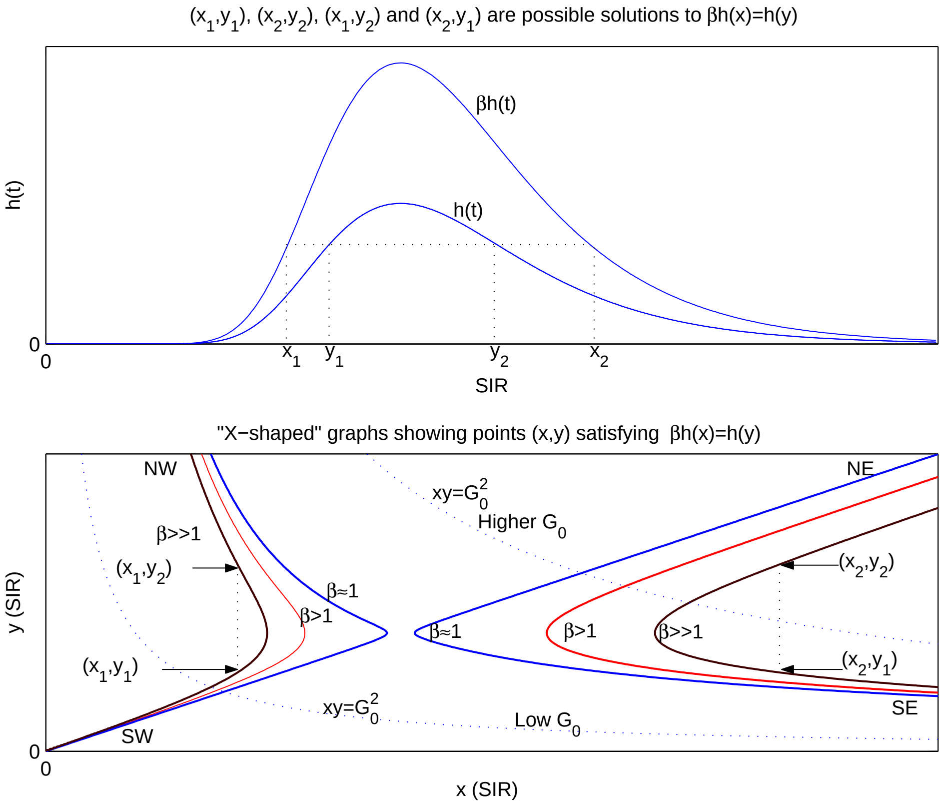
- γ_0 is unique positive solution to $xf'(x) = f(x)$. (see figure)
- It is required that $G_1 = \gamma_0/\sqrt{\beta} \geq G_0$
- Second order conditions indicate this solution is always a “saddle point”
- This allocation is ‘fair’ : both users enjoy same *weighted* throughput

CTM: “Unfair” Boundary Solution

- Set $G_2 = G_0$ ("important" terminal operates at highest available data rate), and $\mu_1 = 0$ ($G_1 \geq G_0$)
- In order for this solution to FONOC to exist, the SIR of the ordinary terminal must be γ_0 (as before), and that of the “important” terminal must be a solution to $x^2 f'(x)/f'(\gamma_0) = G_0^2/\beta$.
- $x^2 f'(x)/f'(\gamma_0)$ has a "bell-shaped" graph. Thus, if G_0^2/β is "too large", $x^2 f'(x)/f'(\gamma_0) = G_0^2/\beta$ has no solutions. Otherwise, it has two solution.
- Let δ_0 be the largest of the two values satisfying the preceding equation. All the optimizing values can be determined in terms of δ_0 and γ_0.
- In particular, $G_1 = \gamma_0\delta_0/G_0$ which must exceed G_0. Some analysis indicates that $G_1 \geq G_0$ requires that $\delta_0 \gg \gamma_0$. Thus G_0^2/β must fall substantially below the “peak” of $x^2 f'(x)/f'(\gamma_0)$.
- Second order conditions confirm that if this solution exists, it is a (local) maximizer

CTM: Greedy allocation

- Seek a solution to FONOC by setting $G_1 = G_2 = G_0$ (both terminals operate at maximal data rate)
- With x and y the SIR of, respect., the important and ordinary terminal, FONOC requires $h(y) = \beta h(x)$ with $h(t) := t f'(t)$, plus the constraint $xy = G_0^2$. Plotting all points that satisfy $h(y) = \beta h(x)$ yields an "X-shaped" graph. The intersections of this "X" with the hyperbolic graph of $xy = G_0^2$ may lead to feasible solutions to FONOC.
- Solution(s) always exists but may be a maximizer or a minimizer.
- $\min\{x, y\}$ must be $\geq \gamma_0$ to ensure the Lagrange multipliers have the correct sign (-ive). Additionally, in order to satisfy the second-order conditions, the maximizer must be in the NE "leg" of the X .
- For equally important terminals, FONOC is satisfied with equal-received power, but could yield a maximizer (with G_0 "large") or a minimizer.



With $h(t) := t\,f'(t)$, FONOC requires that $\beta h(x) = h(y)$, which is satisfied by (x_1,y_1), (x_2,y_2), (x_1,y_2), or (x_2,y_1) (top). All such points form an "X-shaped" graph. For low G_0, the hyperbola (constraint) may *only* intersect the SW leg of the X-curve, which leads to a minimum.

CTM: Recap

- Each physical layer has a "preferred" SIR, γ_0, where $f(x)/x$ is max.
- 3 key allocations: (i) "balanced": both terminals operate at γ_0, and achieve equal weighted throughput; (ii) "unfair" : the "important" terminal operates at maximal bit rate, and the other at the SIR, γ_0; and (iii) "greedy" : both terminals operate at maximal bit rate.
- The balanced assignment is always suboptimal. The important terminal should always operate at maximal data rate. Only when G_0 is "large" should both terminals operate at maximal data rate.
- The "greedy" allocation is treacherous : can lead to either a maximum or a minimum, depending upon whether G_0 exceeds certain threshold.
- Greedy and the unfair allocations "complement" each other.
- Extensions of this analysis to consider noise, and many terminals, some of which could be transmitting media content are available

Overall Discussion – 1

- An analytical foundation for wireless resource management has been discussed, and several specific applications given.
- At the core is one or more functions about which all that is known is that their respective graphs are S-curves. No “equation” is used.
- The family of S-curves include (i) “mostly” concave, (ii) “mostly” convex, (iii) “mostly” straight, (iv) and “step” “curves”.
- An S-curve can yield a useful abstraction of (i) the physical layer of a wireless communication system and (ii) the human visual system
- With f an S-curve, the unique point at which $f(x)/x$ is maximized can be easily identified by drawing a tangent, and plays a fundamental role in several applications.
- Several interesting problems involving coding rate, data rate and power allocation can be solved by applying this framework.

Overall Discussion – 2

- The applications discussed fall in 3 categories: (i) single-terminal, and multi-terminal from a (ii) decentralized and (iii) centralized perspective.
- The single-terminal analysis, whether involving data, or cross-layer allocation for images or video, yields clear and specific answers
- A decentralized allocation involving many terminals can be obtained as a "Nash-equilibrium" of a "game", but it is "inefficient".
- "Mechanism design" can lead to decentralized efficiency. A "compensation mechanism" from the economics literature is proposed.
- The centralized throughput maximization, with weights, is undertaken via classical optimization theory. The solutions to the KKT conditions can be described, and their optimality analyzed, on the basis of the "shapes" of graphs arising from the original S-curve.
- Full papers and additional slides about these and other problems can be obtained at http://pages.poly.edu/~vrodri01/research