Inatel
Instituto Nacional de Telecomunicações

Mestrado em Telecomunicações
TP542 – Otimização Convexa
Prof. Dr. Dayan Adionel Guimarães

Caros alunos e alunas,

Ao longo do segundo semestre de 2013 enviei aos participantes do curso TP542 (antes TP524) uma série de comentários sobre o conteúdo das aulas ministradas a cada semana, objetivando auxiliá-los no entendimento dos conceitos. Esses comentários estão agrupados neste documento e para utilizá-los adequadamente basta estudar os itens correspondentes ao conteúdo ministrado em cada conjunto de, no máximo, quatro aulas.

Além dos exercícios do livro texto, cujo grau de complexidade não é baixo (na média), cinco listas de exercícios foram preparadas e já estão disponíveis na página da disciplina. Recomenda-se que elas sejam resolvidas à medida que o curso for progredindo, de forma contínua e não somente nos dias próximos das provas.

Seguem então os comentários:

- Os conceitos sobre Álgebra Linear são necessários para que consigamos ler com fluidez o livro texto, nos atendo mais aos conceitos sobre Otimização Convexa. A dica é que utilizem as vídeo-aulas ou o livro *Introduction to Linear Algebra* do Prof. G. Strang. As vídeo-aulas estão disponíveis para download na página da disciplina.
- Adquiram o costume de acompanhar as vídeo-aulas do Prof. S. Boyd referentes ao curso sobre Otimização Convexa ministrado em Stanford. O material didático que utilizaremos nas aulas de TP542 é o mesmo utilizado pelo Prof. Boyd em Stanford. As vídeo-aulas do Prof. Boyd também estão disponíveis para download na página da disciplina.
- Recomendo que estudem os Capítulos 1 e 2 do livro texto antes da segunda semana de aulas. Adotem a prática de estudar o conteúdo do livro à frente do conteúdo ministrado em sala. Isto facilitará o entendimento das explicações do professor e gerará dúvidas que podem promover úteis discussões em sala. Não se preocupem em entender todos os detalhes, mas tentem marcar as partes que tiverem mais dúvidas para que possamos discutir na primeira oportunidade.
- Com relação ao *least-squares* (Seção 1.2) gostaria de salientar alguns conceitos referentes a álgebra linear: o *least-squares* (LS) é chamado de solução aproximada do sistema de equações sobredeterminado $Ax = b$, com A de ordem $m \times n$, $m \geq n$, embora este sistema possa ter solução exata. Dica: pesquisem sobre "*system of linear equations*" no Wikipédia; entendam o significado de sistema *overdetermined* e *underdetermined*.
- Ainda com relação ao LS, veja que na solução analítica pressupõe-se que A^TA não seja singular (seja "inversível"). Veja na p. 200 do livro do Strang a condição para que A^TA seja "inversível". Tente descobrir o que é feito quando A^TA não pode ser invertida. Dica: pesquisar sobre matriz pseudo-inversa e tentar descobri como este conceito se aplica à solução do LS.
- Também com relação ao LS, estudem o exemplo 1, p. 206 do livro do G. Strang. Trata-se de uma aplicação de regressão linear bem interessante.
- Observem que no problema de iluminação (slide 1-10) o autor chama de *weighted LS* algo diferente daquilo que ele define na página 5 do livro texto.
- Ainda com relação ao problema de iluminação, no slide 1-11 tentem interpretar a função $h(I_k / I_{\text{des}})$ como sendo composta de duas outras: uma exponencial decrescente obtida se fixarmos I_k e variarmos

I_{des}; outra função linear obtida se fixarmos I_{des} e variarmos I_k. O máximo entre estas curvas resulta naquela da figura do slide.

- Procurem entender a transformação do problema na equação (1.6), p. 6, no problema (1.7), p. 7. O entendimento desta transformação justificará o item 4 do slide 1-10, que propõe a solução do problema de iluminação via programação linear.
- Leiam com cuidado o item 1.6 e visitem a lista de notações ao final do livro, após as referências. Sugiro que leiam mais de uma vez para que comecem a se acostumar com a notação. Lembrem-se de que sem familiaridade com a notação é muito difícil ler o texto com fluidez e, portanto, mais difícil ainda será entendê-lo.
- Sempre que possível, usem o arquivo *ShapesMcd15.xmcd* disponível na pasta da disciplina quando quiserem visualizar algum conceito geometricamente, principalmente nessa parte inicial do curso.
- Atentem para o fato de que a diferença entre as combinações afins, convexas e cônicas está simplesmente nas diferentes restrições dos coeficientes.
- A terceira expressão da p. 25 do livro texto contém uma integral que tem associação com uma grandeza física se C for um corpo no espaço tridimensional. Procurem determinar qual é essa grandeza e como ela se encaixa na interpretação da citada expressão.
- No primeiro parágrafo do item 2.2.1 diz-se que b é uma constante que determina o deslocamento do hyperplano em relação à origem. Sim, ela está associada a este deslocamento, mas não é o próprio deslocamento. Notem que $a^T x_0 = b$ e é x_0 o deslocamento (as componentes de x_0 determinam o deslocamento na direção de cada eixo). O arquivo *ShapesMcd15.xmcd* pode auxiliá-los nessa interpretação.
- Na interpretação das Figuras 2.6, 2.7 e 2.8 talvez seja mais fácil sempre considerar que o vetor a parte da origem e não de um ponto pertencente ao hyperplano a que se refere. Obviamente, quando o deslocamento do hyperplano é nulo, o vetor a partindo da origem também partirá do hyperplano.
- A partir da p. 30 do livro texto aparecem vários conceitos que dependem de álgebra linear. Sobre matrizes (semi)definidas positivas, consultem o livro do G. Strang, p. 330. Sobre autovalores e autovetores, consultem o mesmo livro nas páginas 274-279.
- No título do item 2.2.3, sugiro que traduzam como "Bolas e Cones sob Normas Quaisquer". Na notação da expressão logo no início deste item, $\mathbf{R}^{n+1}$ significa que x pertence a $\mathbf{R}^n$ e t pertence a $\mathbf{R}^1 = \mathbf{R}$.
- Na expressão no Exemplo 2.3 do livro texto (p. 31) utiliza-se uma representação matricial aparentemente muito mais complexa que a representação em linguagem de conjuntos. No entanto, para a solução de um problema de otimização a linguagem de conjuntos normalmente é traduzida como nesse exemplo, o que faz mais sentido do ponto de vista computacional.
- Na Figura 2.11, p. 32, a interseção de cinco meio-espaços está associada à desigualdade na expressão (2.3). Imaginando que tal poliedro está em $\mathbf{R}^3$, a restrição de igualdade em (2.3) pode ser interpretada como aquela que restringe os pontos do poliedro a um plano (a folha do livro neste caso) inserido no espaço $\mathbf{R}^3$.
- Lembrem de que, em $\mathbf{R}^2$, o ortante não negativo é o nosso conhecido "primeiro quadrante", incluindo os eixos (veja o exemplo 2.4 do livro texto).
- Não se preocupem com o item *simplexes* na página 32. Num momento mais adequado daremos a devida atenção a ele, se necessário for.
- A Seção 2.3 é de grande importância. Há conjuntos simples que garantidamente são convexos. Há outros conjuntos que são bastante complexos e cuja convexidade pode ser demonstrada sabendo-se que tais conjuntos podem ser decompostos naqueles mais simples. As operações que preservam convexidade é que dizem como conjuntos complexos podem ser decompostos em conjuntos simples. No Capítulo 3 há uma seção similar que tratará de operações que preservam convexidade de funções.
- Para entender a composição da Fig. 2.14 do livro texto, basta fixar o valor de t na correspondente rotina no arquivo *ShapesMcd15.xmcd*, executar o experimento e notar tal composição como a interseção de várias "fatias" (*slabs*) de pontos que satisfazem a expressão (2.10) para um dado t.

- Os vários mapeamentos lineares citados no item 2.3.2 são importantes ferramentas para a verificação de convexidade de conjuntos mais complexos através da composição por outros mais simples e garantidamente convexos. Uma breve consulta no Wikipédia sobre o termo *affine transformations* pode dar uma melhor visão sobre a interpretação destes mapeamentos do ponto de vista geométrico.

- Não se preocupem por hora com os exemplos 2.9 a 2.12. No entanto, notem que argumentos como aqueles lá encontrados podem ser úteis para provar a convexidade de algum conjunto (e, depois, de alguma função) em um artigo científico, por exemplo.

- Sobre *minimum* e *minimal elements*, vejam o livro do Datorro, 2010, pp. 102-103. Notem particularmente que o conceito de *minimal element* torna-se útil quando um conjunto não contém mínimo. Tentem interpretar a Figura 2.17 à luz das duas primeiras expressões do item 2.4.2.

- Não se preocupem por hora com a Seção 2.5. Creio que ela será mais útil ao entendimento de algoritmos de otimização, assunto que não será abordado no curso. De qualquer forma, vale dar uma olhada no conceito de hiperplanos de separação e no item 2.5.2, pois pode ser útil ao entendimento da Seção 2.6.

- Uma interpretação interessante da eq. (2.19) é aquela em que se considera como pontos do cone dual aqueles que formarem um ângulo igual ou menor que 90 graus com qualquer ponto no cone original. Procurem se lembrar daquela regra de determinação dos limites do cone dual por meio de retas que partem da origem, a 90 graus de cada um dos lados opostos do cone original. Vejam novamente o arquivo *ShapesMcd15.xmcd*.

- Atentem para a definição de produto interno entre matrizes (vejam o apêndice A.1.1).

- Uma maneira útil de interpretar uma desigualdade generalizada do tipo $x \succeq 0$ é dizendo que x pertence ao cone de referência para a correspondente desigualdade. Por exemplo, se $x \succeq 0$ sob a desigualdade baseada no ortante não negativo $\mathbf{R}^2_+$, significa que x pertence a este cone (o ortante não negativo). Se $X \succeq 0$ sob a desigualdade baseada no conjunto de matrizes semidefinidas positivas (cone semidefinido positivo), então X pertence a este cone. Notem que $x \succ 0$ (desigualdade estrita, *strict inequality*) significa que x pertence ao "interior" do correspondente cone de referência.

- Estudem com bastante cuidado o último parágrafo da p. 54 do livro texto, onde se tem uma demonstração matemática sobre a caracterização do elemento mínimo segundo um cone dual.

- Tanto na caracterização do elemento *minimal* quanto do elemento *minimum* segundo um cone dual, lembrem-se daquela regra de "varrer" o conjunto S com hiperplanos definidos por (ou ortogonais a) vetores λ, com λ pertencente ao cone dual em questão. Procurem reanalisar as Figuras 2.23, 2.24 e 2.25 à luz desta regra.

- Há vários conceitos muito importantes no texto que descreve o exemplo 2.27, p. 57. Recomendo que estudem com muita atenção este exemplo.

- Na Figura 2.27, observem que a determinação da fronteira eficiente de produção (que contém os pontos ótimos de Pareto) usando a definição de elemento mínimo da página 45 difere daquela determinada utilizando-se da "varredura" do conjunto P com hiperplanos ortogonais a λ pertencente ao cone dual. Se P fosse um conjunto convexo, notem que ambas a regras levariam aos mesmos pontos ótimos de Pareto.

- Com relação à condição de primeira ordem no item 3.1.3, p. 69, façam um breve estudo do tema "*Taylor series*" no Wikipedia para que vocês interpretem de forma mais acurada a expressão (3.2). Deem uma olhada numa animação que aparece na página em questão. Ela dá uma boa visão sobre o conceito de aproximação de uma função via série de Taylor.

- É bom também darem uma olhada no Wikipedia sobre o conceito de "*gradient*". Sua utilização será muito frequente ao longo de todo o curso.

- Na função "*quadratic-over-linear*", p. 72, atentem para a notação $\mathbf{dom}\, f = \mathbf{R}\times\mathbf{R}_{++}$, o que significa que x pertence a $\mathbf{R}$ e y pertence a $\mathbf{R}_{++}$.

- Nas páginas 72 e 73 há alguns exemplos de funções convexas e as correspondentes provas de convexidade. É importante que entendamos ao máximo tais provas, mas o mais importante para os objetivos do curso é sabermos recorrer a tais provas se um dia precisarmos delas para, por exemplo, demonstrar que alguma função é convexa.

- Na página 74 demonstra-se de forma bastante resumida que a função $f(X) = \text{logdet}(X)$ é côncava. Os desenvolvimentos matemáticos intermediários e mesmo os conceitos que são utilizados não são evidenciados. Tais conceitos já foram incorporados ao arquivo *logdetX.xmcd* na pasta da disciplina.
- Mais uma vez gostaria de lembrar vocês que o estudo constante do livro e o uso das vídeo-aulas do Prof. Boyd devem ser considerados como ferramentas indispensáveis para melhor entendimento do conteúdo. Lembro também que há casos em que é necessário consultar conceitos sobre álgebra linear; as vídeo-aulas do Prof. G. Strang podem ser úteis nesses momentos.
- Estamos acostumados com o uso da média aritmética em várias aplicações. Entretanto, vale a pena tomar conhecimento do conceito e aplicações da média geométrica (vejam a p. 78).
- Fiquem atentos para a conexão entre funções e conjuntos convexos via epigrafo (final da p. 75).
- No exemplo 3.4, p. 76 aparece pela primeira vez o Complemento de Schur. Vejam as páginas 650-651 para mais detalhes. Vejam que no exemplo 3.4 utiliza-se a LMI (*linear matrix inequality*) como forma de mostrar a convexidade do epigrafo da função fracionária matricial (*matrix fractional function*). Vejam também o exemplo 2.10, p. 38 para relembrarem o conceito de LMI e suas propriedades com relação a convexidade. Mais adiante veremos que o complemento de Schur é muito utilizado na modelagem de problemas de otimização convexa, particularmente na transformação de restrições de desigualdade não lineares em LMIs, as quais são normalmente melhro operadas pelos *solvers*.
- Na primeira expressão da p. 77 utiliza-se uma representação vetorial para a última expressão da p. 76. Primeiro tentem identificar que a primeira expressão da p. 77 é de fato a representação de um meio-espaço; vejam a eq. (2.2), p. 29. Em seguida, expandam tal expressão de forma a verificarem que realmente ela representa a última expressão da p. 76. Notem a presença do vetor $[\nabla f(x) \quad -1]^T$. É interessante notar que esse vetor será sempre perpendicular ao hyperplano de suporte da função em qualquer ponto da mesma.
- Fiquem atentos para os conceitos relacionados à desigualdade de Jensen. Eles nos serão muito úteis mais à frente.
- Pesquisem (ou relembrem) sobre o conceito de densidade marginal e tentem associá-lo à segunda expressão da p. 79.
- Para entenderem a representação para o **epi**(wf) na terceira expressão da p. 79, consultem o tema "*linear map*" no Wikipedia e notem que a multiplicação de **epi**f pela matriz está escalonando em w o conjunto **epi**f na mesma direção que é escalonada a função f quando multiplicada por w. Tentem reproduzir este escalonamento em um conjunto qualquer no arquivo *ShapesMcd15.xmcd*.
- Não se preocupem por hora com o item "*representation as pointwise supremum of affine functions*" na p. 83.
- Lembrem-se da dica sobre composição (p. 84-86): em vez de "decorar" as regras que levam a funções convexas ou côncavas, talvez seja mais fácil escrever a expressão da segunda-derivada ou Hessiana, conforme o caso, e determinar quais combinações das funções componentes garantem que o resultado seja maior ou igual a zero (para funções convexas).
- Ainda com relação à composição de funções, fiquem atentos para o uso da extensão de funções (*extended-value extention*) nestas composições. Dica: procurem entender o que pretende ilustrar a Fig. 3.7, p. 85.
- Com relação à composição vetorial, consultem regras de diferenciação sempre que necessário. Vejam, por exemplo, o Apêndice A.4, p. 640 do livro texto ou o Apêndice D.2 do livro do Datorro.
- Para sanar eventuais dúvidas sobre a verificação de convexidade da última expressão do slide 3-18, vejam o segundo item do exemplo 3.14, p. 87 do livro texto.
- O exemplo 3.15, p. 88 contém novos conceitos que merecem atenção, pois aparecerão com frequência no nosso curso: já estudamos o complemento de Schur, mas nesse exemplo é interessante verificar que ele não é determinado diretamente segundo a regra mostrada nas páginas 650-651. Vale a pena tentar desenvolver uma regra própria, usando, por exemplo, "rotação" dos blocos da matriz em análise. Outro conceito importante já mencionado algumas vezes é o conceito de pseudo-inversa (vejam o livro do G. Strang, Cap. 7).

- Na p. 90 do livro texto aparece a definição da "medida" de Divergência de Kullback-Leibler. Em caráter informativo, façam uma breve pesquisa e tentem encontrar aplicações para esta "medida". Trata-se de uma medida muito utilizada na prática, tendo inclusive ramificações ligadas à Teoria da Informação.

- A Seção 3.3, que começa na p. 90 do livro texto, embora um pouco abstrata neste momento, será muito útil no estudo do Cap. 5 sobre Dualidade e, mais importante, no estudo de aplicações. Os exemplos 3.23 a 3.28 são um tanto complexos para o momento. Não se preocupem com eles. Não se preocupem também com a subseção 3.3.2.

- Para interpretar a expressão (3.20), p. 99, referente à condição de primeira ordem para quase-convexidade de funções, primeiro comparem esta condição com aquela referente à convexidade de funções, dada pela expressão (3.2), p. 69. Leiam com atenção os últimos dois parágrafos da p. 100 para facilitar esta comparação. Quanto à interpretação geométrica de (3.20), revisitem a expressão (2.2), p. 29, para relembrarem o conceito de meio-espaço. Depois tentem verificar que a expressão (3.20) sugere que todos os pontos em um dado conjunto-subnível estão localizados no meio espaço definido pelo vetor normal $\nabla f(y)$ para $y = x$.

- A Seção 3.4.4 do livro não foi abordada nos slides e nas vídeo-aulas. Entretanto, vale a pena "dar uma olhada" nela por completo.

- Atentem para o item que trata da integração de funções log-côncavas na p. 106 e correspondentes exemplos nas páginas 107 e 108.

- A Seção 3.6 foi abordada de forma superficial na sala de aula. Recomenda-se apenas a recuperação do conceito de convexidade com relação a desigualdades generalizadas, o qual se assemelha muito ao mesmo conceito quando aplicado à convexidade de conjuntos. Uma leitura com atenção desta seção deve ser suficiente.

- Na p. 128 do livro texto, a restrição $\|z - x\|_2 \leq R$ significa que x minimiza $f_0(z)$ nos pontos próximos deste mínimo. Em outras palavras, qualquer valor de z distante de no máximo R do ponto ótimo (local) levará a valores mais elevados ou iguais para a função objetivo.

- No exemplo 4.1, p. 128 é interessante plotarem as funções para melhor fixarem os conceitos que este exemplo pretende ilustrar.

- Na p. 129, o problema de factibilidade escrito na forma (minimize 0, sujeito a ....) pode retornar apenas dois valores: 0 e $+\infty$. O valor 0 ocorrerá quando x for factível e o valor $+\infty$ ocorrerá quando x não for factível. Note que tal problema retorna UM ponto factível que, em princípio, não tem utilidade. No entanto, quando este problema está no contexto do algoritmo da Bissecção, o último ponto factível determinado será o ponto ótimo (ε-subótimo, na verdade).

- Na Seção 4.1.3 do livro texto, p. 130, no item "*change of variables*" diz-se que a imagem da função f deve cobrir o domínio do problema de otimização. Interpretem com cuidado esta afirmação.

- Para melhor entenderem o item "*transformation of objective and constraint functions*", talvez seja interessante revisitarem o conceito de composição de funções.

- Na p. 132, no processo de eliminação de restrições de igualdade, notem no problema de otimização após a equação (4.8) que a restrição em (4.8) torna-se implícita em $x = f(z)$.

- Ainda na p. 132, notem que o item "*eliminating linear equality constraints*" tem relação com o item anterior, onde $Ax = b \Leftrightarrow x = Fz + x_0$ está associado a $h_i(x) = 0 \Leftrightarrow x = \phi(z)$. Notem que há vários conceitos de álgebra linear no que diz respeito à solução de sistemas de equações lineares. Para entendê-los completamente faz-se necessário estudar o Cap. 3 do livro do G. Strang, com mais atenção à Seção 3.4 (p. 144 daquele livro). Tais conceitos serão novamente trazidos à tona na Seção 4.2.4 do livro texto, no contexto de problemas de otimização equivalentes.

- O método de inserção de restrições de igualdade (p. 132) parece estranho numa primeira análise, mas é muito útil em direção a outras modificações em um problema e que o tornarão realmente mais simples.

- No item "*optimizing over some variables*", p. 133, notem após o problema (4.9) que as restrições devem ser independentes para que tal técnica possa ser utilizada.

- No exemplo 4.4, p. 134, mais uma vez utiliza-se o complemento de Schur. Revisitem o Apêndice A.5.5 do livro texto para relembrarem este conceito e para entenderem o exemplo em questão.

- Entendam bem a forma de epigrafo de um problema de otimização na p.134 e também na p. 143. O conceito está por trás de uma das mais usuais regras de "linearização" de um problema de otimização, na qual uma função objetivo qualquer é convertida em uma função objetivo linear. Para entenderem que a função $t \in \mathbf{R}$ é linear em relação às novas variáveis $x \in \mathbf{R}^n$ e t, vejam que se pode escrever $t = e^T[x\ t]^T$, sendo $e \in \mathbf{R}^{n+1}$ um vetor com os primeiros n elementos iguais a 0 e o último igual a 1.

- Na Figura 4.1, para melhor entenderem forma de epigrafo de um problema de otimização, imaginem uma reta passando por um valor qualquer de t acima do mínimo da função. Imaginem agora esta reta deslocando-se para baixo. Notem que o correspondente conjunto sub-nível será progressivamente reduzido até se tornar um único ponto x em que $f_0(x) = t \Rightarrow p^\star = t$.

- A Seção 4.1.4, p. 136 contém uma abordagem bastante interessante e que amplia nossa visão acerca de como os problemas de otimização podem ser manipulados computacionalmente. Sua aplicação, entretanto, está ligada à implementação de algoritmos, assunto fora do escopo do nosso curso.

- Notem, logo no início da Seção 4.2, p. 136, que a definição de um problema de otimização convexa agora se torna mais precisa do que simplesmente dizer que se trata de um problema em que se minimiza uma função objetivo convexa em um conjunto convexo de restrições. Vejamos: na interpretação geométrica do problema (4.1) pode-se dizer que se tem a minimização da função objetivo em um conjunto que é a intersecção dos conjuntos subníveis 0 (0-*sublevel sets*) definidos por f_i, $i = 1...m$, e os conjuntos de soluções de $h_i(x) = 0$, $i = 1...p$. Pois bem, em um problema de otimização convexa tem-se a minimização da função objetivo convexa em um conjunto que é convexo, pois é a intersecção dos conjuntos subníveis 0 definidos por $\{f_i\}$ convexas (que são convexos, portanto) e o conjunto de soluções do sistema linear $Ax = b$, que é um conjunto afim e, portanto, convexo (veja exemplo 2.1 na p. 22). Lembrem-se de que a intersecção de conjuntos convexos gera um conjunto convexo.

- Para a interpretação da Fig. 4.2, p. 139 têm-se ao menos duas opções: a) $-\nabla f_0(x)^T(y - x) \leq 0$ significa que todos os pontos $y \in X$ estão no meio-espaço definido pelo vetor $-\nabla f_0(x)$. b) todos os pontos em X tem produto interno maior que ou igual a 0 com o gradiente $\nabla f_0(x)$, o que também significa que todos os pontos $y \in X$ estão no meio-espaço definido pelo vetor $-\nabla f_0(x)$.

- Ainda com relação à interpretação da condição de optimalidade para f_0 diferenciável (Seção 4.2.3), notem na Fig. 4.4, p. 147, que o hiperplano de suporte se refere ao conjunto de pontos factíveis, ao passo que, na Fig. 3.12, p. 100, onde nos referimos a funções quase-convexas, o hiperplano de suporte se refere a um conjunto subnível: uma diferença sutil, mas extremamente importante para que entendamos os correspondentes conceitos.

- O item "*unconstrained problems*", p. 140, apresenta uma generalização de um conhecido conceito relacionado à derivada nula. Vale a pena tentarem entendê-lo bem, utilizando como auxílio o seguinte argumento: se o problema de otimização não contém restrições, qualquer valor de $z = y - x$ em $\nabla f_0(x)^T z \geq 0$ é válido. Seja por exemplo o vetor $z = -\nabla f_0(x)$. A única forma de termos $-\nabla f_0(x)^T \nabla f_0(x)^T \geq 0$ é fazendo $\nabla f_0(x) = 0$. Esta é uma forma alternativa (diferente daquela apresentada no livro texto) de mostrar que, no ponto ótimo, $\nabla f_0(x)$ deve ser nulo em problemas de otimização convexa sem restrições.

- Os itens "*problems with equality constraints only*" e "*minimization over the nonnegative orthant*" devem ser estudados, mas os conceitos neles contidos provavelmente ficarão mais claros somente quando estudarmos o Capítulo 5. Portanto, não se preocupem muito com eles agora.

- Na p. 145, a expressão $f_0(x) \leq t \Leftrightarrow \phi_t(x) \leq 0$ é essencial para entendermos como são resolvidos problemas de otimização cuja função objetivo é quase-convexa. Primeiro, ao escrever $f_0(x) \leq t$ estamos nos referindo à forma de epigrafo do problema de otimização quase-convexa. Ao escrevermos $f_0(x) \leq t \Leftrightarrow \phi_t(x) \leq 0$ queremos dizer que $f_0(x) \leq t$ somente se a função convexa $\phi_t(x)$, parametrizada por t, atender a $\phi_t(x) \leq 0$. Portanto, adicionamos a restrição $\phi_t(x) \leq 0$ às restrições do problema quase-convexo e solucionamos um problema de factibilidade em que as desigualdades são convexas e as igualdades são lineares (ou seja, trata-se de um problema de otimização convexa). O problema de factibilidade é normalmente resolvido utilizando-se o método da bissecção (veja slides 4-15 e 4-16 e o final da p. 145 do livro texto). A dificuldade pode aparecer na determinação da função $\phi_t(x)$. Vejam um exemplo em que este procedimento é simples, no slide 4-15.

- Ainda com relação à expressão $f_0(x) \leq t \Leftrightarrow \phi_t(x) \leq 0$, notem que se x é factível para $\phi_t(x) \leq 0$, teremos $f_0(x) \leq t$, o que implica que $t \geq p^\star$. Por outro lado, se x não é factível, t estará abaixo de $f_0(x)$, o que significa que $t \leq p^{\star}$. Estas observações podem ser úteis à interpretação da operação do algoritmo da bissecção (ver slide 4-16).

- A Seção 4.3 aborda a primeira classe de problemas de otimização estudada no nosso curso e, portanto, merece muita atenção ao ser estudada. Além disso, problemas de otimização linear (ou programação linear) são largamente utilizados na prática, mesmo quando a função objetivo não é linear (revisitem a representação de um problema de otimização na forma de epigrafo).
- No inicio da Seção 4.3.2, notem que a solução do problema de otimização quase-convexa pode ser realizada via método da bissecção (revisite os conceitos correspondentes na p. 145).
- Na transformação de um problema de otimização linear-fracionário em um problema linear utiliza-se o conceito de homogeneização, sobre o qual se recomenda pesquisar e estudar um pouco, apenas com nível de profundidade suficiente para se entender tal transformação. Assistam o início da vídeo-aula #6 do S. Boyd (embora isto não vá ajudar muito, creio). Não se preocupem por hora com o exemplo 4.7, referente ao último item desta seção (*generalized linear-fractional programming*).
- No problema (4.34), p. 152 do livro texto pressupõe-se que a matriz P seja semidefinida positiva. Na vídeo-aula #6 o prof. Boyd comenta que se esta condição não for atendida o problema pode tornar-se "*NP-hard*". É importante que façam uma breve pesquisa sobre o significado deste termo no contexto de complexidade computacional e tentem justificar porque o problema (4.34) tornar-se-ia *NP-hard* se $P \prec 0$.
- Na Fig. 4.5, observem que, quase naturalmente, a solução ótima pode se encontrar fora de um vértice do poliedro que representa os pontos factíveis.
- Na p. 153 do livro texto mais uma vez aparece o conceito de pseudo-inversa no contexto da solução do problema dos mínimos quadrados (*least squares*). Revisitem este conceito.
- O assunto "*robust linear programming*" é de grande utilidade prática. Recomenda-se que o correspondente item na p. 157 do livro texto seja estudado com a máxima profundidade.
- Para melhor entenderem a razão do nome SOCP para o problema (4.36), p. 156 do livro texto, revisitem o conceito de cone de segunda ordem (p. 31) e comparem com as restrições de desigualdade de (4.36). Outras denominações para problemas de otimização são dadas de forma análoga (tendo como referência o conjunto associado às restrições ou à função objetivo).
- O assunto "*linear programming with random constraints*" está relacionado ao assunto do item anterior e também é de grande utilidade prática. Recomenda-se que o correspondente item nas pp. 157-158 do livro texto seja estudado com a máxima profundidade. Procurem entender todas as manipulações matemáticas, principalmente aquelas que usam conceitos de probabilidade.
- Não se preocupem por hora com o tópico "*minimal surface*" na p. 159 do livro texto.
- A Seção 4.5 trata da otimização geométrica (GP, *geometric programming*). Esta família de problemas de otimização encontra muitas aplicações em telecomunicações, o que sugere que devemos estudar o assunto com o máximo empenho. Percebam a quantidade enorme de variações que podem ser obtidas com as funções monomiais (ou monômios) e as funções posinomiais (ou posinômios).
- Sugiro que em vez de estudar o problema "*design of a cantilever beam*", que trata de uma aplicação da GP em mecânica, se estude o exemplo da Seção 4.1, p. 80 do artigo "*A Tutorial on Geometric Programming*", disponível na pasta de artigos na página de TP542, pois se trata de uma aplicação em telecomunicações.
- Se vocês já estudaram algo referente a Cadeias de Markov, façam uma pesquisa procurando aplicações para a teoria de Perron-Frobenius (ver p. 165 do livro texto) no equilíbrio de matrizes de probabilidade de transição (matrizes de Markov). Para ajudar, assistam a vídeo-aula #24 do Prof. G. Strang sobre Álgebra Linear. Caso não tenham estudado Cadeias de Markov, pesquisem sobre outra aplicação da teoria de Perron-Frobenius em alguma área de interesse e que esteja relacionada com otimização convexa.
- O início da Seção 4.6, p. 167 do livro texto prepara para os tópicos seguintes no que se refere às desigualdades generalizadas. Aqui a recomendação é simples: revisem o conceito de função K-convexa a partir da p. 109, Subseção 3.6.2.
- Para melhor entenderem o problema (4.49), p. 168, comparem-no com o problema (4.27), p. 146.

- Para melhor entenderem as duas expressões após o problema (4.49), p. 168, comparem-nas com os problemas (4.28) e (4.19), p. 147.
- Para melhor entenderem o problema (4.51), p. 168, comparem-no com o problema (4.28), p. 147. Revisitem o conceito de produto interno entre matrizes na p. 634.
- Com relação a grande parte dos problemas nas Subseções 4.6.2 e 4.6.3, revisitem o conceito que diz que os autovalores de uma matriz diagonal são os próprios elementos desta matriz e que tal matriz é semidefinida positiva se todos os autovalores são maiores ou iguais a zero.
- Para melhor entenderem a modelagem descrita pelo primeiro problema SOCP da subseção 4.6.3, consultem a p. 156 do livro texto, Subseção 4.4.2. Talvez seja bom consultar o conteúdo da p. 43 também.
- Não é necessário estudarem os itens "*moment problems*", "*bounding portfolio risk with incomplete covariance information*" e "*fastest mixing Markov chain on a graph*" nas páginas 170-174.
- É importante estudarem com atenção redobrada a Subseção 4.7.1, p. 174, pois há nela conceitos importantíssimos para que se compreenda melhor a otimização vetorial e a otimização multicritério.
- Quando estiverem estudando as Subseções 4.7.3 e 4.7.4, vale a pena revisitarem os conceitos correspondentes nas páginas 54-58.
- Lembrem-se que problemas de otimização vetorial e multicritério são muito comuns na prática, o que reforça a necessidade de darmos grande atenção ao tema. Vejam que interessantes são as interpretações referentes à escalarização de problemas multicritério, pp. 183-184.
- Este é o momento para começarem a estudar o manual do software CVX disponível na pasta de TP542. Basta desconsiderar aquelas partes sobre Dualidade, assunto que será iniciado em breve. OBS: o quanto antes se acostumarem com a linguagem do CVX melhor. Vale lembrar que o trabalho final utilizará a ferramenta.
- O assunto *dualidade* é um dos mais importantes do curso, principalmente do ponto de vista de sua aplicabilidade prática. Algumas justificativas para esta importância, as quais serão mais bem entendidas ao longo do estudo do Capítulo 5 do livro texto, são:

  1. o assunto servirá para desmistificar os conceitos sobre os *multiplicadores de Lagrange*;
  2. problemas de otimização duais são sempre convexos;
  3. a solução de problemas duais é muito útil na determinação de limitantes inferiores para a solução de problemas não-convexos de difícil solução;
  4. pode-se ter a solução de problemas duais como alternativa para a solução de problemas de otimização convexa;
  5. pode-se ter a solução de problemas duais como alternativa para a solução de alguns (poucos) problemas de otimização não-convexa de difícil solução;
  6. a solução de problemas duais sempre serve como algum tipo de certificação sobre a optimalidade da solução de problemas de otimização convexos;
  7. a solução de problemas duais sempre serve como algum tipo de certificado sobre o grau de proximidade entre a solução do problema dual e a solução de um problema original complexo ou não convexo;
  8. a solução de problemas duais pode servir como critério de parada em algoritmos de otimização.

- Existe um relação bastante estreita entre a *dualidade de Lagrange* e o processo de regularização para a solução de problemas de otimização multicritério. Esta relação pode ser vista por meio das "funções de irritabilidade" mencionadas ao final da vídeo-aula #7 do Prof. Boyd, ou por meio do estudo da Subseção 5.3.3, pp. 236-237 do livro texto.
- Uma interpretação interessante para os multiplicadores de Lagrange na expressão do Lagrangeano (*Lagrangean*) na p. 215 pode ser: $\{\lambda_i\}$ é o "preço" que se paga para satisfazer mais ou menos ao objetivo e às restrições de desigualdade; de maneira análoga, $\{\nu_i\}$ pode ser interpretado como o "preço" que se paga por se aceitar maiores ou menores desvios com relação às restrições de igualdade.
- Atentem para o fato de que a função dual de Lagrange é côncava independente do problema de otimização primal ser ou não ser convexo.

- No item "*standard form LP*", p. 219, para entenderem porque a função $g(\lambda, \nu)$ é côncava, recorram à dica do Prof. Boyd no início da vídeo-aula #8, quando ele fala sobre este problema.
- Segundo o Prof. Boyd, o problema "*two-way partitioning*", p. 219 tem muitas aplicações práticas e é um ótimo exemplo de problema combinatorial de solução extremamente difícil quando o número de elementos em x é grande. Vale a pena entendê-lo bem, principalmente porque mais adiante no livro texto ele será transformado em um SDP cuja solução é relativamente simples (via CVX, por exemplo), mesmo para n muito grande. Esta transformação é bastante recente nos estudos sobre otimização.
- Para melhor entenderem o slide 5-6, vejam a p. 221 do livro texto, principalmente a eq. (5.11). Vejam também o slide 5-8 (eu colocaria este slide antes do 5-6, para facilitar).
- As qualificações de restrição (*constraint qualifications*) estabelecem condições (além da análise de convexidade) sob as quais a dualidade forte se aplica. São, *per si*, um extenso ramo no estudo de otimização. A qualificação de Slater é apenas uma das possíveis formas de qualificação de restrição, embora muito útil no contexto de problemas de otimização convexa, garantindo, sob certas circunstâncias, $p^\star = d^\star$ (dualidade forte).
- Notem que a dualidade fraca $d^\star \leq p^\star$ é válida mesmo se o problema original (primal) não for convexo. Se o problema primal é ilimitado abaixo (*unbounded below*), então $p^\star = -\infty$, o que necessariamente implica em $d^\star = -\infty$, significando que o problema dual não é factível. Por outro lado, se o problema dual é ilimitado acima (*unbounded above*), então $d^\star = \infty$, o que necessariamente leva a $p^\star = \infty$, significando que o problema primal não é factível.
- Fiquem atentos aos comentários no início da p. 227 do livro texto (antes da Subseção 5.2.4).
- Na página 229 do livro texto, no item "*a nonconvex quadratic problem with strong duality*" tem-se um exemplo de um problema de otimização não convexa com dualidade forte. Na maior parte das vezes não é simples provar esta dualidade forte e no livro tem-se um exemplo. Vale a pena ao menos ler este item para terem uma noção da complexidade da questão, ficando atentos para o parágrafo final da p. 229 onde se tem uma importante conclusão.
- Não se preocupem por hora com a Subseção 5.2.5, pp. 230-231. Idem para a Seção 5.3, pp. 232-236, exceto a Subseção 5.3.3, p. 236, que merece ser estudada com atenção. Não se preocupem por hora com a Seção 5.4, pp. 237-241. Retomem na Seção 5.5, p. 241. Ela trata de um assunto de grande importância prática e teórica.
- Para mais bem entenderem a folga complementar (*complementary slackness*) na última expressão da p. 242, usem o fato de que, se $\inf[g(x)] \leq g(x)$ para qualquer x, isto também é válido para $x = x^\star$. Isto justifica a passagem da segunda para a terceira linha da expressão. As demais justificativas estão bem claras no livro, após tal expressão.
- As condições de KKT têm grande importância prática. Além de poderem representar uma forma alternativa para a solução analítica de problemas de otimização convexa em alguns casos, em qualquer problema de otimização com função objetivo e restrições diferenciáveis e com dualidade forte, qualquer conjunto de pontos ótimos para os problemas primal e dual tem que satisfazer as condições de KKT.
- Quando o problema de otimização é convexo, as condições de KKT são também suficientes para a igualdade entre os pontos ótimos dos problemas primal e dual. Em outras palavras, se os pontos x, λ e ν satisfazem as condições de KKT, obrigatoriamente serão ótimos. Isto indica que a solução de problemas de otimização convexa pode ser vista como um problema de solução do sistema de equações formado pelas condições de KKT.
- Quando o problema de otimização é convexo, com função objetivo e restrições diferenciáveis, e satisfaz a condição de Slater de factibilidade estrita (*strictly feasibility*), as condições de KKT são suficientes e necessárias para optimalidade. Em outras palavras, a condição de Slater implica em *gap* de dualidade nulo e que o ponto ótimo dual é atingível. Então x será ótimo para o problema primal se e somente se houver (λ, ν) ótimos para o problema dual que, juntamente com x, satisfaçam às condições de KKT.
- A justificativa de que o gradiente do Lagrangeano deve ser nulo em uma das condições de KKT (veja p. 243 do livro texto) pode ser dada à luz da primeira cadeia de expressões no início da Subseção 5.5.2 (*complementary slackness*). Lá se pode concluir que $x^\star$ minimiza o Lagrangeano $L(x, \lambda^\star, \nu^\star)$,

pois nesta cadeia de expressões tudo deve ser satisfeito com a igualdade. Então $\inf L(x, \lambda^\star, \nu^\star) = f_0(x^\star)$ significa que $x^\star$ é um minimizador de $L(x, \lambda^\star, \nu^\star)$, o que exige $\nabla L(x^\star, \lambda^\star, \nu^\star) = 0$.

- Estudem o exemplo 5.2, p. 245 com a ajuda da solução do problema 8 da 5ª lista de exercícios.
- Não se preocupem por hora com a Subseção 5.5.4, pois ela trata da interpretação das condições de KKT num contexto de mecânica.
- A solução de um problema primal via problema dual é bastante interessante, principalmente naqueles casos em que a solução do problema dual pode ser realizada analiticamente ou de forma menos complexa. Faz-se da seguinte maneira: encontra-se a solução do problema dual, o que leva aos pontos ótimos $\lambda^\star$ e $\nu^\star$. Se a dualidade forte se aplica, o ponto ótimo primal $x^\star$ é um minimizador de $L(x, \lambda^\star, \nu^\star)$. Então, resolvendo-se o problema de minimização de $L(x, \lambda^\star, \nu^\star)$ encontra-se a solução que, se for factível para o problema primal, fornecerá o ponto ótimo $x^\star$ para este problema. Se a solução para o mínimo de $L(x, \lambda^\star, \nu^\star)$ existir, mas não for factível para o problema primal, então se diz que o ótimo primal não é atingível (como no caso de min. $1/x$: o mínimo é 0, mas não é atingível por nenhum x finito).
- Para entenderem a expressão (5.57), p. 250 do livro texto, na demonstração logo em seguida usem o fato que se $\lambda_i \geq 0$ e $f_i(x) \leq u_i$, então $\Sigma \lambda_i f_i(x) \leq \Sigma \lambda_i u_i$.
- A Seção 5.6, p. 249 é de grande importância teórica e prática. A análise de sensibilidade é baseada na análise dos multiplicadores de Lagrange (soluções do problema dual), sendo que estes estão disponíveis “gratuitamente” nos *solvers*, por definição. Em outras palavras, tornou-se um padrão implementar *solvers* que resolvem simultaneamente os problemas primal e dual. Por este ângulo, de fato os multiplicadores de Lagrange nos são fornecidos gratuitamente.
- Na expressão (5.56), p. 249 pode-se interpretar variações em u_i como afrouxamentos ($u_i > 0$) ou a atribuição de mais rigor ($u_i < 0$) nas restrições de desigualdade. Já as variações em v_i podem ser interpretadas como desvios ou deslocamentos nos valores alvo (*set points*) de restrições de igualdade. Em ambos os casos tais variações ou alterações propositais normalmente são seguidas de um novo processo de otimização (reprojeto), por meio de uma nova solução para o problema de otimização com as novas (e alteradas) restrições. O quanto tais restrições podem ser alteradas pode ser conhecido justamente com a análise de sensibilidade apresentada na Seção 5.6.
- Com relação ao assunto do slide 5-22, suponham que $f_1(x^\star) = 0$, o que força $\lambda_1^\star > 0$ devido à folga complementar. Suponham que $\lambda_1^\star = 0.001$ (supostamente um valor pequeno para o problema em questão). Isto significa que se pode redimensionar o projeto promovendo alguma alteração na correspondente restrição. Devido ao fato de $\lambda_1^\star$ ser pequeno, saberemos que tal alteração não produzirá grandes variações em $p^\star$, o valor ótimo do problema de otimização primal. Em outras palavras, $\lambda^\star$ e $\nu^\star$ pequenos indicam que pouca variação ocorrerá em $p^\star$ se as restrições forem alteradas ou perturbadas. Estas são reais e importantes interpretações sobre o significado prático dos multiplicadores de Lagrange.
- Complementando esta questão de sensibilidade, pela expressão (5.58), p. 252 do livro texto podemos ver que, de fato, λ e ν representam a taxa de variação em $p^\star$ devido às variações das restrições.
- A Seção 5.7, p. 253, tem o principal objetivo de mostrar que simples transformações na descrição (modelagem) de um problema primal podem resultar em problemas duais completamente diferentes. Isto pode ser útil para que, com simples transformações no problema primal, se produza um problema dual de solução mais simples. Por outro lado, a variedade de problemas duais que podem ser obtidos a partir de simples transformações nos primais pode trazer diferentes modelagens que, por sua vez, podem se “encaixar” em uma grande variedade de aplicações. Isto pode ser de fato muito útil do ponto de vista prático.
- Não é preciso estudar a Seção 5.8 do livro texto, pp. 258-264. O próprio Prof. Boyd não a abordou em suas vídeo-aulas. O mesmo comentário vale para a Subseção 5.9.4. No entanto, o Prof. Boyd afirma que o Teorema de Alternativas é um assunto de grande importância, pois trata da factibilidade de problemas duais. Em havendo tempo, vale a pena estudar este tema.
- Vejam que os problemas de aproximação por norma do Cap. 6 são mais úteis quando $m > n$, ou seja, em sistemas sobredeterminados, em que o número de equações é maior que o número de incógnitas no sistema $Ax \sim b$. Nesse caso não há solução exata, apenas aproximada.
- Na interpretação do problema de aproximação por norma para o caso do modelo de medida linear $y = Ax + v$, a solução mais plausível é aquela que melhor aproxima $y - Ax$ de v, mas isso é mais plausível quando o ruído tem baixa intensidade, ou seja, quando $\|v\|$ é pequena. Isso implica que a

solução mais plausível é aquela que minimiza $\| Ax - y \|$. Vejam que isso exige o conhecimento *a priori* de que $\|v\|$ é pequena.

- A função pênalti *deadzone-linear* (p. 295) é facilmente implementada no CVX como max(0, $|u|-a$).
- A função pênalti *log barrier* apresenta um grande desafio para implementação no CVX. Para se ter uma ideia, consulte o exemplo "*penalty_comp_cvx.m*" referente ao Cap. 6 no diretório do CVX.
- Percebam que a minimização do pênalti total com funções pênalti $\phi(u) = |u|$ e $\phi(u) = u^2$ corresponde à minimização de $\| Ax - b \|$, respectivamente nas normas ℓ_1 e ℓ_2.
- A função pênalti de Huber, muito utilizada para regressão linear com pouca sensibilidade aos *outliers*, embora seja de difícil implementação direta no CVX, tem uma função (*atom*) já incorporada à biblioteca do CVX o que, portanto, facilita em muito o seu uso na prática.
- Vejam as interessantes observações sobre sensibilidade aos *outliers* no primeiro parágrafo da p. 299.
- Para facilitar a interpretação do exemplo na p. 296, usem a interpretação da "*marginal irritation function*", definida como a derivada da função pênalti. Quando analisarem os diferentes pênaltis aplicados aos resíduos, pensem em pênaltis relativos, não absolutos.
- Usando a "*marginal irritation function*", interpretem o papel da função pênalti da Figura 6.3, p. 298.
- Deem atenção especial à aproximação por minimização da norma ℓ_1. Problemas de minimização de norma ℓ_1 estão intimamente ligados com um importante ramo que tem muitas aplicações na nossa área: *compressed sensing*.
- Vejam com atenção o segundo item do tópico "*norm ball constraint*". p. 302.
- Os problemas de minimização de norma descritos a partir da p. 302 (Seção 6.2), ao contrário dos problemas de aproximação por norma, são interessantes quando $m < n$, ou seja, em sistemas subdeterminados nos quais o número de equações é menor que o número de incógnitas. Aqui há infinitas soluções e, às vezes, *uma única solução esparsa*. Essa solução esparsa é, acredito, uma propriedade intrínseca de alguns sistemas de equações lineares subdeterminados (vale pesquisar sobre isso). A solução de problemas de minimização de norma ℓ_1 leva, em grande parte dos casos, a soluções esparsas para o problema de minimização de norma (enquanto levava a resíduos esparsos nos problemas de aproximação por norma; atenção para essa sutil diferença).
- Vejam na equação (6.6) uma variação do problema de minimização do pênalti total associado aos resíduos $r = Ax - b$, agora associado aos próprios valores de x. Compare essa equação com a (6.2).
- Estudem com especial atenção o item "*sparse solution via least* ℓ_1*-norm*" na p. 304.
- Recomenda-se atenção especial ao estudo do exemplo 6.3, p. 307, pois se trata de uma aplicação muito ligada à nossa área e a DSP genericamente. Aqui a típica desconvolução em processamento de sinais é enriquecida com restrições sobre a forma de onda de entrada que se deseja dimensionar. Pode-se interpretar esse exemplo como um problema de pré-distorção que não poderia ser resolvido com simples desconvolução.
- As funções objetivo do problema multicritério do exemplo 6.3, p. 307 podem ser respectivamente interpretadas como:

    1. Erro quadrático médio (*mean square error*) entre o sinal desejado e o obtido. Esta é uma medida usual de discrepância, pois como vimos mais recentemente, o processo de aproximação por norma referente ao *least squares* produz um estimador de máxima verossimilhança (MV). Portanto, quando se minimiza o MSE, implicitamente busca-se a solução de MV para o problema em pauta.
    2. Potência média (*average power*), associada à intensidade do sinal.
    3. Primeira diferença quadrática média (*mean square first difference*), que é uma medida análoga à derivada e que, portanto, está associada ao grau de "suavização" do sinal desejado.

- O processo *smoothing regularization* (p. 307) ao qual o exemplo 6.3 se refere contém informações bem interessantes sobre a realização das operações de convolução, primeira derivada e segunda derivada, assim como tem o exemplo *quadratic smoothing* (p. 312). Vejam que tais operações podem ser efetuadas pela multiplicação da uma matriz D pelo vetor operado x. Deem especial atenção à construção dessa matriz D em cada caso, devido à importância prática. Vejam o código CVX do exemplo 6.3 e aqueles referentes às Figuras 6.8-6.11.

- No processo de reconstrução de sinais por meio da função de suavização do tipo variação total (*total variation smoothing*), veja que a minimização da diferença de primeira ordem (ou primeira diferença), por ser feita na norma ℓ_1 produzirá solução esparsa, ou seja, com grande quantidade de zeros (ou quase zeros). Mas aí vem a pergunta: que tipo de sinal tem primeira diferença esparsa? Resposta: um sinal do tipo constante por partes (*piecewise constant*). Tal sinal tem derivada nula nas partes constantes e não nulas nas transições (vejam, por exemplo, o gráfico superior da Figura 6.14, p. 317, tendendo a ser um sinal *piecewise constant*).

- Com relação à aproximação robusta (p. 318), percebam que na abordagem "nominal" e "estocástica" tem-se solução analítica e que na abordagem de "pior caso" a minimização do supremo de uma função quadrática em um intervalo pode ser reescrita como a minimização do máximo entre os valores da função nos pontos extremos do intervalo (p. 320).

- No final do Cap. 6, a Seção 6.5 não foi abordada em sala nem nos slides e vídeo-aulas. No entanto recomenda-se sua leitura, com atenção especial para as Subseções 6.5.4 e 6.5.5, pois podem ter aplicações interessantes na nossa área.

- Já no assunto sobre estimação estatística, percebam que a maximização da função de log-verossimilhança parametrizada por x é em relação à variável x e não a y (p. 351). Há casos em que, dependendo do parâmetro, uma função $\ln p_x(y)$ pode não ser log-côncava em x para determinado y.

- No modelo de estimação de x a partir de medidas em um modelo linear $y_i = a_i^T x + v_i$, a PDF da medida y_i segue a distribuição da PDF do ruído, apenas deslocando-se a sua média. Em outras palavras, se $p(z)$ é a PDF do ruído, $p(y_i) = p(z)$ para $z = y_i - a_i^T x$.

- Ainda com relação ao modelo linear supracitado, percebam que se o ruído for Gaussiano, o vetor x de máxima verossimilhança é o mesmo que aquele encontrado via minimização de $\| Ax - y \|_2$. Veja que se $A = I$ e x representa um vetor-sinal transmitido em um sistema de comunicação digital, sendo y o vetor recebido ($y = x +$ ruído), o vetor x de máxima verossimilhança será aquele que minimiza $\| x - y \|_2$, ou seja, será o vetor sinal mais próximo de y em termos de distância Euclidiana. Essa solução coincide com a estrutura do receptor de máxima verossimilhança em canal com ruído Gaussiano no contexto de transmissão digital. Se as variâncias de ruído são diferentes nas m medidas, o problema se tornará um *weighted-LS* (verifique isso como exercício).

- Ainda com relação ao modelo de estimação de x a partir de medidas em um modelo linear $y_i = a_i^T x + v_i$, se o ruído for Laplaceano, o vetor x de máxima verossimilhança é o mesmo que aquele encontrado via minimização de $\| Ax - y \|_1$. Vejam que se um ruído impulsivo puder ser modelado como um ruído Laplaceano, o correspondente receptor em um sistema de comunicação digital será aquele que minimiza $\| x - y \|_1$, ou seja, será o vetor sinal mais próximo de y em termos da norma ℓ_1.

- A Subseção 7.1.2 (*MAP estimation*) não foi abordada formalmente e nem o será. No entanto, trata-se de uma generalização importante da estimação de máxima verossimilhança. Recomenda-se, portanto, o seu estudo como parte do conteúdo do curso.

- A título de curiosidade, no contexto de estimação paramétrica, quando o parâmetro estimado é um inteiro (por exemplo, as médias condicionadas à transmissão de 0s e 1s em um sistema de comunicação digital), é comum denominar o problema como problema de detecção em vez de problema de estimação. Assim, estimar corretamente um parâmetro equivale a detectá-lo corretamente.

- A Seção 7.2 não foi abordada em sala e não consta dos slides. Entretanto, por se tratar de um assunto comum na nossa área, recomendo que o estudem.

- A Seção 7.3 trata do projeto de detectores ótimos no contexto de teste de hipóteses. Logo no início da seção conceituam-se as distribuições de probabilidade p e q. Para facilitar a visualização da aplicação prática, admitam que tais distribuições sejam versões discretas (e adequadamente normalizadas) de funções densidade de probabilidade da variável de decisão em análise sob as possíveis hipóteses.

- É de suma importância interpretar a matriz T, pois é ela a variável do problema de otimização que busca projetar o detector ótimo. Portanto, o projeto a que se refere tal problema consiste em determinar T. Como exemplo, considere $T = [1\ 1\ 1\ 0\ 0.2\ 0;\ 0\ 0\ 0\ 1\ 0.8\ 1]$, na sintaxe do Matlab. Esta matriz se refere a um detector binário (T tem duas linhas, cada uma referente a cada hipótese) que tem como entrada uma variável X com n = 4 valores (que é o número de colunas em T), por exemplo, 1, 2, 3 e 4. Associando a primeira linha de T à hipótese H_1 e a segunda à hipótese H_2, pode-se dizer: se X = 1, 2 ou 3, decide-se pela hipótese H_1; se X = 4 ou 6 decide-se pela hipótese H_2; se X = 5 decide-se pela hipótese H_1 com probabilidade 0.2 e pela hipótese H_2 com probabilidade 0.8. Este é a operação do detector projetado a partir da matriz T. Neste exemplo trata-se de um *detector aleatório*; quando T tem somente 0s e 1s tem-se um *detector determinístico*.

- Dediquem especial atenção à Seção 7.3.4. Ela contém uma série de conceitos importantes sobre teste de hipóteses.

- A matriz W logo após a eq. (7.12) atribui pesos (custos) às decisões, sendo nulos os custos referentes às decisões corretas. Neste sentido, o problema (7.13) busca minimizar o custo médio total de se decidir erroneamente.

- Mais adiante, na subseção "*MAP and ML detectors*", vejam que se faz $W_{ij} = q_j$. Assim, se um determinado q_j tem valor alto, ele penalizará mais o correspondente erro, fazendo com que tal erro seja mais reduzido ao longo da execução do algoritmo de otimização. Analisando por outro ângulo, o detector projetado tenderá a priorizar a decisão pelas hipóteses mais prováveis, ação coerente com o uso de probabilidades *a priori* como forma de "refinar" as decisões, como acontece no critério MAP.

- No teste binário de hipóteses, a razão de verossimilhança p_k/q_k permite decidir pela hipótese H_1 se $p_k/q_k > 1$ e por H_2 caso contrário. Nesse caso não há nenhuma tendência em dar preferência para uma ou outra hipótese. Assim se faz no critério ML. Já quando se compara a razão de verossimilhança com um limiar $\lambda \neq 1$, interpreta-se como uma situação em que há tendência em dar mais ou menos prioridade à decisão por uma das hipóteses. O limiar, é claro, será função das probabilidades *a priori*. Assim se faz no critério MAP.

- A Seção 7.3.6 não foi abordada em sala, mas é bastante interessante em termos de aplicação prática na nossa área. Portanto, recomenda-se estuda-la.

- Não se preocupem por hora com a Seção 7.4 (embora haja conceitos interessantes nela, especialmente o exemplo na Seção 7.4.3).

- A apresentação formal do nosso curso termina aqui. Utilizaremos o tempo restante para estudo preparatório para a segunda prova e para a elaboração do trabalho final. Há outros temas interessantes nos capítulos 7 e 8, particularmente *experiment design* e *classification*. A Parte III do livro (capítulos 9 em diante) não foi e nem será tratada como assunto do curso.

- Ainda que não tenhamos estudado todo o conteúdo do livro texto, o que foi estudado já é bastante extenso e complexo. Um bom resumo do conteúdo pode ser útil nesse momento. Recomenda-se que os tutoriais I e II do H. Hindi [6][7] sejam estudados, excetuando, caso assim desejem, a parte sobre os métodos de ponto interior (*interior point methods*), os quais se referem a algoritmos para a solução de problemas de otimização. No entanto, essa parte traz interessantes informações acerca do uso da dualidade como critério de parada dos algoritmos. Os tutoriais supracitados estão disponíveis na pasta de artigos na página da disciplina.

Dayan Adionel Guimarães, Janeiro de 2014