

Universidade de Brasília

Faculdade de Economia, Administração, Contabilidade e

Ciência da informação e Documentação

Departamento de Economia

**OTIMALIDADE E DUALIDADE NA LIQUIDAÇÃO DE
PAGAMENTOS INTERBANCÁRIOS**

MÁRIO RUBEM DO COUTTO BASTOS

BRASÍLIA – DF

2009

Universidade de Brasília

Faculdade de Economia, Administração, Contabilidade e

Ciência da informação e Documentação

Departamento de Economia

**OTIMALIDADE E DUALIDADE NA LIQUIDAÇÃO DE PAGAMENTOS
INTERBANCÁRIOS**

por

MÁRIO RUBEM DO COUTTO BASTOS

Dissertação apresentada ao Departamento de Economia da Universidade de Brasília, como parte dos requisitos para obtenção do grau de

MESTRE EM ECONOMIA

08 de maio de 2009

Comissão Examinadora:

Prof. Dr. Rodrigo Andrés de Souza Peñaloza – Orientador

Prof. Dr. Daniel Oliveira Cajueiro – Membro

Dra. Adriana Soares Sales – Membro

A minha esposa, Elaine Licio
pelo carinho e compreensão durante
todos os dias de nosso casamento.

AGRADECIMENTOS

À Deus acima de tudo.

À minha esposa Elaine pelos conselhos que me tranqüilizaram durante os longos períodos de provas; pela paciência em ler, ouvir e melhorar textos e apresentações de assuntos em que não tinha interesse; por discutir questões ideológicas que se contrapunham aos meus interesses de estudo, mas que me incentivaram a aprofundar na teoria econômica; e, principalmente, por existir e me fazer uma pessoa muito feliz.

Aos professores da Unb pela dedicação e perseverança na arte de lecionar, destacando o meu orientador e amigo Rodrigo Peñaloza que, mesmo antes de iniciar este projeto, me inspirava com suas aulas que combinavam modelagem matemática e filosofia econômica.

A todos os colegas do Departamento de Operações Bancárias e Sistema de Pagamentos (DEBAN) que, de forma direta ou indireta, me introduziram o tema dessa dissertação e me fizeram amadurecer conceitos que foram intensamente utilizados.

Aos colegas do mestrado pela atenção e carinho que transformaram as aulas em um prazeroso ambiente de aprendizagem, dividindo as conquistas com fraternidade e descontraindo o ambiente nos intervalos dos estudos. Destaco uma dedicação especial aos amigos Bruno, Carlos Felipe, Rafael e Paulo pela camaradagem, a qual espero manter para sempre.

RESUMO

Essa dissertação analisa um modelo híbrido do Sistema de Pagamentos onde as transferências são acumuladas em filas que permitem a liquidação pelo valor líquido da posição multilateral do participante. Quando os recursos no sistema são insuficientes para liquidar todos os pagamentos existentes, a solução ótima depende da exclusão de alguns pagamentos do processo de liquidação. Esse problema, conhecido na literatura como Banking Clearing Problem (BCP), é NP completo sendo a solução inviável quando existe um número significativo de bancos e pagamentos. A modelagem do BCP por programação linear oferece uma alternativa, possibilitando priorizar a liquidação dos pagamentos que possuem maior capacidade de gerar novas liquidações e permitindo a elaboração de algoritmos de busca da solução ótima mais eficientes.

Palavras-Chave: Sistema de Pagamentos, Risco Sistêmico, Programação Linear, Dualidade, LBTR, LDL, Banco Central.

ABSTRACT

This dissertation analyses a hybrid model of Payment Settlement System where transfers are accumulated in queues allowing a multilateral net settlement. When funds are not enough to settle all existing payments, the optimal solution requires the exclusion of some payments from the process. This problem, known as Banking Clearing Problem (BCP), is NP complete and its solution is time demanding, which renders it not viable in the case of a large number of banks or payments. BCP modeling by linear programming offers an alternative, using the shadow-price of each payment to identify the potential to generate new settlements and allowing the development of more efficient algorithms in the search of optimal solutions.

Keywords: Payment System, Systemic Risk, Linear Program, Duality, RTGS, DNS, Central Banks.

SUMÁRIO

1. Introdução: O Sistema de Pagamentos
2. Revisão da Literatura
3. O Modelo de Resolução de Filas
4. O Algoritmo de Liquidação
5. Performance do algoritmo
6. Conclusão

1. Introdução: O Sistema de Pagamentos

Os bancos e outras instituições financeiras possuem demandas, por conta própria ou de clientes, de transferências de recursos¹ que são realizados ao longo do dia no mercado interbancário. Essas demandas são motivadas por razões diversas, como por exemplo, pelo uso de cheques, cartões, transferências eletrônicas ou compra de ativos financeiros. O Sistema de Pagamentos é a infra-estrutura que possibilita a liquidação desses pagamentos entre as instituições.

Em decorrência das inovações financeiras, da integração de mercados financeiros e da globalização, os valores transferidos nas operações interbancárias aumentaram substancialmente nas últimas décadas, suscitando estudos e políticas quanto ao risco sistêmico induzido por contágio, ou seja, a propagação do problema de uma instituição para as demais afetando a estabilidade do sistema financeiro como um todo. Esta nova percepção de risco contrastou o ambiente de baixa regulamentação que predominava desde o início dos anos 70 e trouxe à vista a necessidade de equilibrar os ganhos de eficiência oriundos da desregulamentação vis-à-vis o aumento risco no setor financeiro. Após a crise na bolsa americana em 1987, os bancos centrais passaram a mais atenção com a gestão do risco, sendo promovidos diversos avanços desde então². A década de 90 também trouxe avanços significativos no desenvolvimento da tecnologia de informação, com o aumento a velocidade e a redução do custo de transferência de informação.

Nesse cenário, o Sistema de Pagamentos foi reestruturado em diversos países, tendo a administração de risco como epígrafe. No que tange à transferência de valores, principalmente de grandes valores, o Sistema de Pagamentos existente na maioria dos países, até o final dos anos 80, era o **Sistema de Liquidação Líquida** (ou **Sistema de Liquidação Diferida Líquida - LDL**). Tal sistema se caracteriza por dois pontos principais: pela separação entre as etapas de processamento e liquidação efetiva dos pagamentos, e pela compensação entre os valores a pagar e a receber. Ao longo do dia

¹ A transferência dos recursos ocorre no momento da liquidação final, irrevogável e incondicional das obrigações de uma instituição com moeda de banco central, ou seja, com o lançamento contábil nas contas de ambas as instituições na autoridade monetária.

² Como exemplos de avanços tem-se a implementação do seguro depósito em diversos países, as diversas reestruturações do sistema de pagamentos, a regulação prudencial tendo como marco o acordo da Basileia etc.

todos os pagamentos são processados, ocorrendo o armazenamento dos valores a pagar e a receber. Numa etapa posterior ocorre a apuração do saldo líquido de cada instituição – podendo ser calculada com base bilateral ou multilateral – e, então, a transferência dos recursos entre as partes. Assim, haverá um conjunto de bancos com posição credora e outro conjunto com posição devedora de recursos. A principal vantagem desse sistema é o baixo custo de liquidação, pois o valor apurado após compensação é, em geral, significativamente menor que a soma dos pagamentos realizados ao longo do dia, reduzindo o custo de oportunidade no carregamento de liquidez para a instituição. Embora eficiente na gestão da liquidez, o Sistema LDL tradicionalmente não limitava o saldo negativo de uma instituição com as demais, permitindo que, ao acumular posição devedora muito grande ao longo do dia, todas as demais instituições fiquem reféns da eventual inadimplência de um ou mais bancos na liquidação.

A partir dos anos 90, vários países passaram a adotar o **Sistema de Liquidação Bruta** (ou **Sistema de Liquidação Bruta em Tempo Real - LBTR**). Nesse sistema, as transferências são liquidadas individualmente, sem compensações e em tempo contínuo³, com as ordens de transferências de valores processadas e liquidadas simultaneamente. Ressalta-se, também, que as transferências ocorrem de forma irrevogável e requerendo que a parte emitente tenha saldo suficiente para a transferência do recurso⁴.

Este novo sistema trouxe maior segurança cabendo ressaltar **três** fatores principais. Ao processar e liquidar as transferências simultaneamente e em tempo real, o sistema retira o risco da insolvência do banco emissor, ou seja, a contraparte fica limitada a realizar operações somente com o saldo em recursos disponível no momento, que aumenta ou diminui conforme o banco receba ou efetue pagamentos ao longo do dia. Ao contrário, no Sistema de Liquidação Líquida, um banco com saldo esperado credor pode aplicar os recursos antes da liquidação efetiva dos valores – em geral no final do dia após a apuração de todas as transferências. Contudo, quando uma instituição aceita

³ A liquidação no sistema bruto pode ocorrer também em intervalos de tempo discretos, quando todas as transferências de pagamentos são acumuladas e processadas individualmente. Todavia, com o avanço tecnológico, atualmente o sistema com liquidação bruta em tempo real é o mais comum e será o único considerado neste trabalho.

⁴ Caso a instituição não tenha saldo para efetuar a transferência, alguns Sistemas LBTR permitem que a transferência fique armazenada em fila – enfileiramento – aguardando o saldo suficiente para liquidação ou que utilizem os recursos previamente disponibilizados – daylight overdraft account.

ordens de pagamento ao longo do dia sem a garantia da solvência no momento da liquidação, fica implícita a concessão crédito à contraparte, tornando-a vulnerável no caso de inadimplência desta no final do período⁵. O risco de liquidação nesse sistema ocorre, então, quando algum participante tem fundo insuficiente para liquidar suas obrigações no momento devido. Neste caso, diferentes procedimentos podem ser adotados, como o reprocessamento (*unwinding*), excluindo algumas ordens de pagamento ou o participante inadimplente. Esse processo pode gerar um efeito cascata, se estendendo a outros bancos que tenham posição líquida credora com a instituição inadimplente e posição líquida devedora com outras instituições.

Também importante, a necessidade da existência de saldo suficiente⁶ de recursos na conta do participante emitente para efetivação de transferências trouxe significativa redução no risco de crédito incorrido pelos bancos centrais. No caso da inadimplência de uma instituição no sistema LDL, bancos centrais podem ser instados a assumir a posição devedora de um banco inadimplente, assumindo com recursos públicos o custo privado para evitar o risco sistêmico. Embora traga evidente custo político, esse cenário é justificável quando o custo social de uma crise sistêmica for maior do que o custo privado da falha inicial.

Finalmente, o Sistema de Liquidação Bruta possibilita que a liquidação financeira ocorra de forma simultânea à transferência de custódia em operações de compra e venda de ativos financeiros – modelo de entrega contra pagamento. Esse é o caso onde um banco, por interesse próprio ou de terceiros, é obrigado a ter saldo suficiente em conta para comprar títulos financeiros. Ao contrário, no Sistema de Liquidação Líquida, a transferência da custódia é efetivada antes da liquidação financeira e o vendedor do ativo ficava vulnerável ao risco de insolvência do banco comprador.

O Sistema LBTR, todavia, ao reduzir o risco sistêmico, aumenta também a necessidade de saldo em disponibilidades para efetuar transferências ao longo do dia. O custo de manutenção da liquidez pode ser equilibrado, em alguns casos, com os recursos provenientes de outros bancos, gerando o incentivo para que as instituições atrasem o

⁵ A diferença entre o sistema líquido e o crédito interbancário é que neste o risco é precificado e há possibilidade de racionar o crédito a uma IF em decorrência de melhores informações. Naquele, porém, a linha de crédito é automática e garantida.

⁶ Alguns bancos centrais, a exemplo do FED, facultam ao banco o direito de operar com saldo negativo em suas contas, embora o saldo máximo seja limitado a um valor previamente informado ao banco.

envio das transferências. Em alguns sistemas é permitido o enfileiramento das ordens de transferências quando o banco não dispõe de recursos para as transferências, ficando estas armazenadas em fila aguardando o saldo suficiente para liquidação. O Sistema LBTR pode, então, se caracterizar pelo aparecimento de filas com diversos bancos enfileirando as transferências. Tal cenário, denominado de *gridlock* (BIS 1993), ocorre quando um grande número de transferências fica pendente de liquidação pela insuficiência de saldo de recursos significativamente menor, e é decorrente do fato das transferências terem que ser liquidadas uma por vez.

Com a substituição do sistema LDL pelo LBTR nos bancos centrais em diversos países, torna-se evidente a opção desses pela segurança em prejuízo da eficiência. Por um lado, o Sistema LDL permite aos bancos maior economia de liquidez e, conseqüentemente, maior disponibilidade para investimentos, mas gera maior risco ao sistema financeiro. Por outro lado, o Sistema LBTR é mais seguro, porém é mais ávido por liquidez – exigindo que as transferências ocorram individualmente pelo valor bruto e sem a possibilidade de saldo negativo em conta – e de maior complexidade de gerenciamento, pois os bancos passam a gerir seus recursos em tempo real.

Em suma, se no LDL as instituições ajustavam sua liquidez de forma a minimizar o seu custo de oportunidade, no LBTR elas têm a necessidade de reter maior liquidez e de gerenciar seu saldo em tempo real.

Esta dissertação tem por objetivo analisar o sistema de pagamentos focando nas transferências de grandes valores, propondo uma metodologia que permita resolver o problema de *gridlock* no Sistema LBTR de forma eficiente. O algoritmo proposto também terá aplicação na resolução do *unwinding* no Sistema LDL quando o banco central buscar uma solução que maximize o valor dos pagamentos sujeito à restrição de liquidez presente no sistema. Este relatório se encontra dividido em cinco seções além dessa introdução. A seção 2 apresenta uma revisão na literatura comparando as vantagens e problemas entre os sistemas LDL e LBTR. A seção 3 apresenta o modelo e a seção 4 mostra sua aplicação através da elaboração de um algoritmo usado na resolução das filas. A seção 5 analisa os resultados obtidos na implementação do algoritmo e a seção 6 conclui o relatório.

2. Revisão da Literatura

Esta seção apresenta, sucintamente, alguns modelos que analisam o comportamento dos agentes nos sistemas de pagamentos.

Freixas e Parigi (1998) exploram o *trade-off* entre eficiência e segurança no LDL e LBTR. Em seu artigo, apresentam um modelo que compara ambos em cenários sem e com incerteza no momento da liquidação.

Em resumo, o paper avalia os equilíbrios decorrentes do comportamento estratégico dos clientes de bancos. No modelo, os bancos possuem o monopólio da tecnologia de investimentos, pertencem aos correntistas, são neutros ao risco e permitem os agentes escolherem o momento do resgate. Os Sistemas LBTR e LDL são modelados seguindo as descrições apresentadas na seção anterior. Adiciona-se, contudo, que no LBTR o banco pode vender com deságio – perda de rentabilidade – uma fração de seus investimentos para atender as demandas por saques e por transferências. Já no LDL há obrigação contratual de uma instituição receber as transferências da outra, ocorrendo, de forma compulsória, uma concessão implícita de crédito entre as instituições. Considera-se também que a liquidação de uma instituição ocorre no final do dia, possibilitando aos agentes agirem estrategicamente ao longo do dia quando dotados de novas informações ou crenças sobre a saúde financeira de seu banco.

Nesse cenário, o sistema LBTR é menos eficiente que o LDL, pois ao liquidar uma parte dos investimentos no valor integral das transferências, reduz a lucratividade dos correntistas. Em situações de estresse financeiro, quando a percepção sobre um determinado banco torna-se negativa, o Sistema LBTR mostra-se mais estável, pois os bancos estão isolados um dos outros. Ao contrário, o Sistema LDL permite o contágio entre as instituições e todos os equilíbrios encontrados – corrida bancária no banco “saudável” ou garantindo sobrevivência ao banco ineficiente – são ineficientes⁷.

⁷ Dois equilíbrios são verificados. No primeiro os agentes do banco afetado pelas informações transferem os recursos para o outro banco. Esse comportamento é explicado pela intenção desses agentes em obter “carona” com os bons resultados da outra instituição. No segundo, todos os agentes de ambos os bancos sacam os depósitos. Esse equilíbrio tem por base a hipótese de conhecimento comum (*common knowledge*), onde os agentes do banco saudável se antecipam ao comportamento dos agentes do outro banco. Estes, então, também sacam seus recursos, pois não há vantagem na transferência dos depósitos.

Em suma, Freixas e Parigi reforçam o *trade-off* entre eficiência e segurança, destacando que o Sistema LBTR é preferível quando o custo de manter um banco ineficiente é alto, quando as transferências interbancárias são pequenas e quando o risco inerente à atividade bancária for alto.

Lester, Millard e Willison (2006) apresentam um modelo os agentes econômicos na sociedade determinam entre os sistemas LDL e o LBTR aquele que maximiza seu bem-estar. O Sistema LDL apresenta um risco potencial aos vendedores quando, ao vender um produto, são obrigados a conceder um “crédito” ao banco, gerando um custo esperado na venda. De outra forma, o Sistema LBTR é mais seguro, a transferência financeira ocorre concomitantemente a “venda” da mercadoria, mas exige um custo devido à tecnologia empregada pelos bancos para administrar a liquidez em tempo real. Este custo é repassado pelos bancos aos clientes através da diminuição da rentabilidade de seus investimentos bancários. O modelo mostra que o equilíbrio vai depender do custo de liquidação no Sistema LBTR, sendo este preferível quando o custo for baixo.

Angelini (1998) modela o comportamento estratégico de bancos na gestão de recursos para pagamentos onde liquidez intradia é disponibilizada pelo banco central sem necessidade de colateral e com custo proporcional ao montante do empréstimo. No modelo, o atraso dos pagamentos gera insatisfação do cliente que gostaria de ver o volume transferido o mais rápido possível. O equilíbrio encontrado no modelo, contudo, não é ótimo, pois o custo do crédito intradia induz os bancos a atrasar os pagamentos, tomando carona na liquidez de outros bancos ao invés de obter o crédito custoso. Estes atrasos resultam em externalidades negativas, afetando a qualidade da informação disponível para o gerenciamento do caixa das contrapartes e, conseqüentemente, podendo gerar perda de rentabilidade das instituições ao induzir níveis sub-ótimos no saldo em reservas das instituições no final do dia. Como sugestão de políticas, o autor que os empréstimos de liquidez devem ser suficientemente baratos – menor que o custo da insatisfação dos clientes aos atrasos dos pagamentos –, mas com colaterais para evitar o risco moral. Ele também sugere que os bancos poderiam ser induzidos a pagar mais cedo se as taxas cobradas fossem mais caras no decorrer do dia, penalizando os pagamentos tardios.

Bech e Garrat (2003) analisam de forma mais ampla o comportamento dos agentes na gerência da liquidez em um sistema LBTR. Os autores analisam uma economia com dois bancos simétricos, com o Sistema de Pagamentos gerenciado pelo BC e com

provimento de liquidez com uso de colateral, com custo proporcional ao montante tomado para saque a descoberto e sem custo para saque a descoberto. Em resumo, o artigo trata de um jogo bayesiano onde cada banco tem informação privada de sua demanda por transferência, que pode ser requerida no período da manhã ou no da tarde com probabilidades definidas e de conhecimento comum. Os bancos agem de forma estratégica procurando minimizar os custos de oportunidade do aporte de garantias e o custo de reputação com os clientes decorrente do adiamento da transferência dos fundos para o período subsequente⁸. Ressalta-se que os recursos recebidos pela contraparte podem ser usados para pagamentos ou “emprestados” para outra instituição caso não sejam requeridos e haja demanda. O comportamento estratégico pode então ser resumido pela possibilidade de postergar ou não a liquidação de uma ordem de pagamento recebida pela manhã.

Para o Sistema LBTR com crédito mediante colaterais, o equilíbrio em estratégias puras dependerá da relação entre os custos de oportunidade (C) da manutenção do colateral e os custos de reputação (D) com os clientes. Se $C > D$ então o equilíbrio será pela efetivação das transferências no período da manhã e se $C < D$ então ambos os bancos postergarão essa ação para a tarde. Contudo, somente o equilíbrio com ambos os bancos fazendo os pagamentos na parte da manhã é eficiente, ou seja, gera o menor custo *ex-ante*.

Bech e Soramäki (2001) propõem um modelo com um mecanismo de resolução de filas – gridlock – no Sistema LBTR. Os autores identificam a possibilidade de reduzir o volume de pagamentos enfileirados pela liquidação simultânea de pagamentos ao invés da liquidação individual dos pagamentos, respeitando-se, porém, que em nenhum momento o saldo de recursos de qualquer banco poderá ser negativo (restrição de liquidez) e que as liquidações deverão ocorrer segundo um ordenamento pré-definido pela instituição (restrição seqüencial). Em suma, trata-se de um mecanismo de otimização de pagamentos armazenados em uma central e sujeito a duas restrições. O algoritmo proposto no modelo reduz a complexidade computacional do Problema de Liquidação de Pagamentos – *Banking Clearing Problem* – cuja otimização do fluxo de pagamentos é restringida somente pela restrição de liquidez, e possibilita a resolução do algoritmo sem aproximações.

⁸ O artigo considera que todas as demandas por transferências devem ser atendidas até o final do dia. Assim somente as demandas no período da manhã poderão ser procrastinadas.

Penalozza (2002) também propõe um modelo que propõe a maximização do fluxo de pagamentos em fila em Sistemas LBTR como um problema de programação linear onde a solução primal do problema decorre das restrições de enfileiramento dos pagamentos, fragmentação das ordens de pagamento e créditos intradia mediante contrapartida em colaterais. O modelo também considera o problema de programação linear dual, obtendo os preços-sombras associados às restrições do problema, sendo estas: restrição de pagamentos (a diferença entre recebimentos e pagamentos é limitada inferiormente pelo saldo inicial em recursos no banco central); restrição de consistência (todos os pagamentos são efetivados); restrição de crédito (o volume de crédito ofertado pelo banco central é limitado pelo valor do colateral utilizado, previamente precificado).

3. O Modelo de Resolução de Filas

A introdução apresenta, de forma genérica, as características dos dois principais tipos de sistemas de pagamentos interbancários para transferências de grande valor – o Sistema LDL e o Sistema LBTR. Ambos os sistemas se diferenciam pelo momento da liquidação dos pagamentos, sendo que no Sistema LDL ocorre o acúmulo das ordens de pagamentos ao longo de um determinado período para, em momento posterior, haver a liquidação financeira pelo valor líquido. No sistema LBTR, contudo, o processamento e liquidação das transferências ocorrem simultaneamente e pelo valor bruto das transferências, tornando-o mais intenso no uso dos recursos para pagamentos. Embora eficiente na gestão dos recursos, o Sistema LDL é mais vulnerável ao risco sistêmico, pois a bancarota de um banco⁹ com posição devedora antes da liquidação dos pagamentos pode resultar em contágio às contrapartes, e, por extensão, a todo o sistema. Assim, bancos centrais têm adotado o Sistema LBTR para transferências de grande valor na maioria dos países desenvolvidos.

O Sistema LBTR pode permitir o enfileiramento das ordens de pagamentos quando não houver recursos suficientes para sua liquidação. Este mecanismo possibilita aos bancos aproveitarem a liquidez oriunda dos demais, onde as ordens de pagamentos permanecem enfileiradas numa central aguardando a entrada de recursos de outra instituição, sendo liquidadas, em geral, pelo o critério FIFO. Este processo de enfileiramento trás, contudo, um risco potencial ao sistema quando os bancos, ao procurarem tomar carona na liquidez dos demais, enfileiram grande quantidade de pagamentos com a expectativa de receber liquidez suficiente para cursar suas obrigações. Tal comportamento é motivado pelo interesse privado de diminuir o custo de oportunidade de reter recursos líquidos próprios de menor rentabilidade e, se dominante entre os participantes, trás consigo o risco da suspensão de grande parte dos pagamentos, podendo causar risco de contágio a outras instituições.

Bech e Soramäki (2001) diferenciam o enfileiramento em duas diferentes situações. A primeira, denominada *gridlock* pelos autores, ocorre quando o sistema apresenta liquidez suficiente para eliminar um subconjunto dos pagamentos enfileirados, caso seja

⁹ Denomina-se banco todas as instituições que puderem ter saldo em recursos no banco central e autonomia para movimentar estes recursos em interesse próprio ou de terceiros, sem intermédio de outra instituição.

permitida a liquidação simultânea entre pagamentos e recebimentos. A segunda, denominada *deadlock*, ocorre se não há no sistema liquidez suficiente para eliminar qualquer subconjunto, ainda que com a compensação entre os pagamentos. A diferenciação entre os enfileiramentos é importante, pois o enfileiramento por *gridlock* apresenta potencial para otimização dos pagamentos, ao contrário das filas em *deadlock* que requerem a ação do banco central como prestamista de última instância para poder haver a liquidação dos pagamentos pendentes.

Esta seção analisa a rede formada pelos pagamentos interbancários em uma situação de *gridlock*, onde há um conjunto de pagamentos enfileirados sendo todos de **igual** prioridade para seus bancos emissores¹⁰. A prioridade desses pagamentos, contudo, não é indiferente sob a perspectiva do banco central. Quando um pagamento é liquidado, ele transfere liquidez a outro banco que passa a poder efetuar novos pagamentos. Dessa forma, pagamentos possuem prioridades diferentes para o banco central que esteja determinado a maximizar o valor das liquidações sob restrição de liquidez. Colocando de outra forma, dado que todos os pagamentos não poderão ser liquidados conjuntamente, um banco central precisa de uma regra que determine a prioridade dos mesmos, de forma a manter enfileirado o menor valor possível. Para esse fim, utiliza-se o framework proposto por Penaloza (2002), modelando o problema do *gridlock* como um problema de programação linear e utilizando o preço-sombra dos pagamentos para estabelecer sua prioridade.

O modelo também possibilita uma solução para Sistemas LDL em caso de *unwinding* – quando uma instituição não é capaz de honrar sua posição devedora e há exclusão de algumas ou todas as ordens de pagamento para novo resultado multilateral. A resolução de filas no Sistema LBTR se assemelha à liquidação em Sistema LDL quando se permite a compensação entre pagamentos e recebimentos¹¹. Trata-se, em ambos os casos, da liquidação pelo saldo após as compensações entre os pagamentos e

¹⁰ Denomina-se banco emissor ou cedente o banco que emite a ordem de pagamento para outro, autorizando o banco central a debitar seu saldo em Reservas Bancárias e creditar a conta de outro banco.

¹¹ A resolução de filas com compensação, assim como a solução para sistemas LDL com inadimplência de pelo menos uma instituição com posição devedora, apresenta algumas diferenças operacionais. No Sistema LDL é necessário que os bancos com resultado multilateral negativo efetuem as transferências dos valores para uma central de pagamentos no Banco Central. Na resolução de filas, no entanto, a otimização é feita com a compensação dos pagamentos no sistema, e, portanto, sem necessidade de adicionar liquidez ao mesmo.

os recebimentos acumulados em determinado período. O modelo em análise, então, busca uma solução factível a ambos os modelos.

Considere, então, um vetor $B = \{1, 2, \dots, n\}$ de bancos que realizam pagamentos ao longo do dia e um vetor $S^0 = \{S_1^0, S_2^0, \dots, S_n^0\}$ definido como o saldo de recursos no início do dia de cada instituição.

Todas as transferências são efetuadas através de um **sistema de mensageria**, onde o banco envia ao banco central uma ordem requerendo o débito contábil em sua conta e o crédito na conta de outra instituição. Ao longo do dia, cada instituição recebe a demanda por transferência e escolhe o momento de efetivá-la, ou seja, enviar as mensagens (ou ordens de pagamento) ao banco central. Se a instituição não possui os recursos necessários, ela pode postergar o envio das ordens, ou enviá-las para que fiquem enfileiradas em uma central até obter saldo necessário. Seja $\{x_{ij}(t) : i \in B, j \in B \setminus \{i\}\}$ o conjunto de mensagens enviadas pelo banco i ao banco j e enfileiradas no momento t , sendo $x_{ijk}(t)$ a k -ésima mensagem enfileirada entre os dois bancos. Para exemplificar, considere uma transferência entre dois bancos. Então o vetor $x_{12} = \{x_{121}; x_{122}; \dots; x_{12K}\}$ mostra que há K transferências enfileiradas do “banco 1” para o “banco 2”. Cabe ressaltar que as movimentações ocorridas dentro de uma mesma instituição – as transferências entre clientes dentro de um mesmo banco – não requerem uso de recursos e por isso não são consideradas nesse modelo.

Uma vez enfileirada, cada transferência pode ser liquidada pelo valor integral ou parcial informado na ordem enviada. O fracionamento do valor das transferências tem a vantagem de aumentar a eficiência com que a liquidez do sistema é empregada, aumentando a velocidade com que permeia o sistema e, conseqüentemente, o número de transações processadas. Tal procedimento reduz a probabilidade de *gridlocks*, embora traga desafios legais e operacionais quando, por exemplo, alguma instituição torna-se insolvente e é incapaz de realizar o pagamento integral até o momento de encerramento das transferências. O fracionamento também pode ser obtido ao se limitar o valor máximo para cada transferência, evitando que valores muito grandes congestionem o sistema¹².

¹² Um exemplo é o Sistema Fedwire, cujo valor máximo para transferência de recursos é limitado a \$50 milhões.

Para capturar a dinâmica do processo de liquidação dos pagamentos, considere $v_{ijk}(t)$ a fração do pagamento k feito pelo banco i ao banco j até o momento t . Esta variável, definida doravante como **função de liquidação**, terá valor igual a um ($v_{ijk}(t) = 1$) quando o pagamento for plenamente efetuado em t . Da mesma forma, $v_{ijk}(t) = 0$ se o pagamento estiver integralmente pendente. A imagem da função liquidação dependerá da arquitetura do sistema, sendo $v_{ijk}(t) \in [0, 1]$ se houver fracionamento das ordens de pagamento, ou $v_{ijk}(t) \in \{0, 1\}$ se as transferências só forem liquidadas pelo valor integral.

Para o gestor do sistema de pagamentos, a solução eficiente de *gridlock* melhora o funcionamento do sistema e reduz o risco de liquidez, efetuando mais rapidamente a liquidação dos pagamentos. Sua importância está não somente na redução dos custos de oportunidade dos bancos, que poderão gerir de forma mais eficiente seus recursos, mas também na menor dependência desses aos instrumentos de crédito intradia. Assim, deve o banco central procurar maximizar o volume de pagamentos, ou seja:

$$(3.1) \quad \text{Max} \sum_{i \in B} \sum_{j \in B} \sum_{k=1}^{N^*} x_{i,j,k} v_{i,j,k}$$

Por simplificação didática, considera-se que todos os bancos transferem N^* mensagens para cada outro. Economicamente, seria como se N^* fosse o maior número de ordens de transferência enfileiradas que uma instituição transfere para outra e se um banco qualquer (banco A) tiver $M < N^*$ mensagens enfileiradas a outro (banco B), então $x_{ABM+1} = x_{ABM+2} = \dots = x_{ABN^*} = 0$.

O problema de maximização está sujeito à restrição de liquidez, em que nenhum banco pode manter saldo negativo em sua conta em qualquer momento. Esta restrição pode ser colocada analiticamente como:

$$(3.2) \quad \sum_{j \in B} \sum_{k=1}^{N^*} x_{ijk}(t) v_{ijk}(t) \leq S_o^i + \sum_{j \in B} \sum_{k=1}^{N^*} x_{jik}(t) v_{jik}(t_k) + \left[\sum_{j \in B} \sum_{k=1}^{N^*} x_{jik}(\bar{t} < t) - \sum_{j \in B} \sum_{k=1}^{N^*} x_{ijk}(\bar{t} < t) \right]$$

A desigualdade acima mostra que o total de pagamentos efetuados por um banco em um dado momento t não pode ser maior que a soma: do saldo do início do dia; do montante proveniente de outras instituições, mas ainda pendente de liquidação; e do valor líquido recebido até esse momento.

O problema de otimização dos pagamentos em fila torna-se:

$$(3.3) \quad \text{Max} \quad \sum_{i \in B} \sum_{j \in B} \sum_{k=1}^{N^*} x_{ijk}(t) v_{ijk}(t)$$

$$\text{s.a:} \quad \sum_{j \in B} \sum_{k=1}^{N^*} x_{ijk}(t) v_{ijk}(t) - \sum_{j \in B} \sum_{k=1}^{N^*} x_{jik}(t) v_{jik}(t) \leq S_t^i$$

$$0 \leq v_{ijk}, v_{jik} \leq 1$$

Onde S_t^i é o saldo da instituição do instante t , ou seja:

$$S_t^i = S_o^i + \left[\sum_{j \in B} \sum_{k=1}^{N^*} x_{jik}(\bar{t} < t) - \sum_{j \in B} \sum_{k=1}^{N^*} x_{ijk}(\bar{t} < t) \right]$$

Para colocar o problema na notação matricial, considere $\mathbf{x}_i(t)$ o vetor de todos os pagamentos emitidos pelo banco i que estão em fila. Os pagamentos devem ser agrupados pelo banco a que se destinam. Por simplicidade, considera-se que o banco i efetue N^* pagamentos para todos os demais bancos, ou seja, $i \in B, j \in B \setminus \{i\}$. Assim:

$$(3.4) \quad \mathbf{x}_{ij}(t) = [x_{ij1}(t), x_{ij2}(t), \dots, x_{ijN^*}(t)]$$

$$\mathbf{x}_i(t) = [x_{i2}(t), x_{i3}(t), \dots, x_{iB}(t)]$$

O vetor de todos os pagamentos pendentes no enfileiramento $\mathbf{x}(t)$ é definido então como:

$$\mathbf{x}(t) = [x_1(t), x_2(t), \dots, x_B(t)]$$

Da mesma forma, o vetor da função liquidação do banco i $\mathbf{v}_i(t)$ e o vetor da função liquidação do sistema $\mathbf{v}(t)$ são definidos como:

$$(3.5) \quad \mathbf{v}_{ij}(t) = [v_{ij1}(t), v_{ij2}(t), \dots, v_{ijN^*}(t)]$$

$$\mathbf{v}_i(t) = [v_{i2}(t), v_{i3}(t), \dots, v_{iB}(t)]$$

$$\mathbf{v}(t) = [v_1(t), v_2(t), \dots, v_B(t)]$$

O produto escalar $\mathbf{x}(t)' \cdot \mathbf{v}(t)$ define a função-objetivo do problema.

Para se obter a condição de restrição, considere então \mathbf{Q}_i uma matriz com dimensão $B \times (B-1)N^*$ e $\mathbf{Q}_{i,l}$ a l -ésima linha da matriz. Conforme já observado, o vetor $x_i(t)$ é constituído pelos elementos $x_{i,j,k}$, sendo i o banco emissor do pagamento, j o banco cujo pagamento se destina e k um índice da mensagem. Então, $\mathbf{Q}_{i,l}$ é definida conforme a regra:

$$(3.6) \quad Q_{i,l} = \begin{cases} 1 * x_i(t) & \text{se } i = l \\ -1 * x_i(t) & \text{se } j = l \\ 0 * x_i(t) & \text{se } j \neq l \end{cases}$$

Exemplificando, considere a situação onde há 4 bancos no sistema interbancário e o banco 1 tem exatamente 2 pagamentos enfileirados para cada outro banco ($k=\{1,2\}$). A matriz Q_I tem a forma:

$$(3.7) \quad Q_I = \begin{vmatrix} x_{121} & x_{122} & x_{131} & x_{132} & x_{141} & x_{142} \\ -x_{121} & -x_{122} & 0 & 0 & 0 & 0 \\ 0 & 0 & -x_{131} & -x_{132} & 0 & 0 \\ 0 & 0 & 0 & 0 & -x_{141} & -x_{142} \end{vmatrix}$$

Finalmente, seja Q a matriz de dimensão $B \times B(B-1)N^*$ definida por:

$$(3.8) \quad Q = [Q_I \mid Q_2 \mid \dots \mid Q_B]$$

A outra restrição impõe limite à função de liquidação e é representada conforme abaixo:

I – matriz identidade com dimensão $B(B-1)N^*$

$\mathbf{1}$ – vetor de 1 com dimensão $B(B-1)N^*$

Verifica-se então que o problema colocado anteriormente pode ser representado em notação matricial conforme:

$$(3.9) \quad \begin{aligned} \text{Max} \quad & \mathbf{x}' \cdot \mathbf{v} \\ \text{sa.:} \quad & \begin{vmatrix} Q \\ I \end{vmatrix} \cdot \mathbf{v} \leq \begin{vmatrix} S \\ \mathbf{1} \end{vmatrix} \\ & \mathbf{v} \geq 0 \end{aligned}$$

onde S é o saldo em recursos do banco no momento da maximização.

O dual do problema de otimização colocado acima pode ser obtido facilmente, sendo apresentado abaixo:

$$(3.10) \quad \begin{aligned} \text{Min} \quad & S' \cdot \lambda + I' \mu \\ \text{sa.:} \quad & \begin{vmatrix} Q^T & I \end{vmatrix} \cdot \begin{vmatrix} \lambda \\ \mu \end{vmatrix} \geq x \\ & \lambda, \mu \geq 0 \end{aligned}$$

O dual é simétrico ao problema anterior, procurando minimizar os recursos necessários para cursar os pagamentos contidos no vetor x . A solução do problema acima apresenta os preços-sombra do saldo em liquidez para pagamentos (λ) – “preço-sombra do saldo” – e da consistência do vetor de liquidação (μ) – “preço-sombra dos pagamentos”. Como as funções-valores do problema primal e dual têm o mesmo valor, os preços-sombras representam a “sensibilidade” do total de pagamentos à variação marginal dos parâmetros S e 1 . Assim, $\lambda_i > 0$ representa a variação no valor liquidado dado o incremento de uma unidade monetária no saldo de recursos.

Observando mais atentamente a restrição de problema dual, observa-se facilmente que esta se traduz na relação abaixo:

$$(3.11) \quad x_{ijk} \lambda_i - x_{ijk} \lambda_j + \mu_{ijk} \geq x_{ijk}$$

Após reorganizar as variáveis, observa-se:

$$(3.12) \quad \mu_{ijk}/x_{ijk} \geq 1 + (\lambda_j - \lambda_i), \quad x_{ijk} > 0$$

Nessa relação, nota-se que o preço-sombra μ_{ijk} apresenta-se dividido pelo valor da transferência. A razão econômica para tal fato está em sua própria natureza, pois sua variação reflete a variação percentual do pagamento. Exemplificando, aumentando o valor do limite de v_{ijk} de 1 para 2 possibilitaria o banco dobrar o valor da transferência – de x_{ijk} para $2x_{ijk}$. Para obter o efeito da variação em unidade monetária é necessário dividir o preço-sombra pelo valor da transferência.

A função objetivo do problema dual tem os parâmetros S e 1 positivos, e, sendo um problema de programação linear, o conjunto restrição é convexo e possui um número finito de vértices no múltiplano – formado pela interseção das restrições. Quando a solução ótima é factível – ou seja, existente e determinada – a função valor é tangente a um dos vértices do conjunto restrição e a relação em (3.11) torna-se uma igualdade. Nesse caso, pode-se obter o preço-sombra dos pagamentos diretamente pelo preço-sombra do saldo.

$$(3.13) \quad \mu_{ijk}/x_{ijk} = 1 + (\lambda_j - \lambda_i), \quad x_{ijk} > 0$$

Caso, porém, o Problema de Liquidação dos Pagamentos ou *Banking Clearing Problem (BCP)* exija solução inteira – $v_{ijk}(t) \in \{0, 1\}$ – a complexidade para encontrar a

solução ótima aumenta. Esse problema é NP completo, o que significa que o tempo necessário para encontrar o resultado ótimo cresce exponencialmente com o tamanho das variáveis envolvidas, tornando-se inviável operacionalmente.

Bech e Soramäki (2001) propuseram a adição de uma restrição seqüencial ao **BCP**, onde os bancos decidem o ordenamento pelo qual ordens são executadas e não há duas ordens de transferência com o mesmo ordenamento. A introdução dessa restrição reduz enormemente a complexidade do problema¹³, mas restringe a qualidade da solução se os bancos forem indiferentes ao ordenamento com que os pagamentos são efetuados. Este trabalho tem por fim analisar o **BCP**, visando a encontrar um conjunto de conceitos que permitam subsidiar algoritmos que permitam uma solução computacional para o problema.

O framework proposto, através da análise da solução ótima do problema por programação linear, permite a organização da rede de pagamentos para que a solução ótima seja obtida. Os bancos que possuem recursos – na conta reserva ou proveniente de outros pagamentos – para liquidar todos os pagamentos terão $\lambda=0$. Ao contrário, quando algum banco não possui recursos suficientes, a restrição de liquidez será “ativa” e $\lambda \in \{1, 2, \dots\}$. Quando um banco com escassez de recursos fica pendente de pagamentos a outro com abundância de recursos, observa-se $\lambda=1$. Economicamente interpreta-se esse valor como o acréscimo na liquidação dos pagamentos decorrente da obtenção de uma unidade extra de valor pelo banco com restrição de recursos.

Dependendo da distribuição dos pagamentos, haverá pagamentos enfileirados “em cascata”, ou seja, pagamentos mantidos em fila e destinados a bancos que também tem pagamentos enfileirados. Para ilustrar, considere três bancos sem recurso e com dois pagamentos em fila, conforme a figura abaixo. O banco A tem um pagamento pendente para o banco B, que também tem outro pagamento em fila para o banco C.

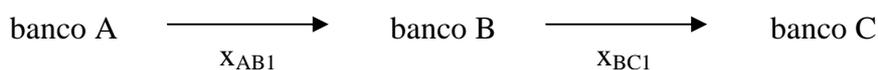


Figura 3.1

¹³ Seja n a soma de todos os pagamentos enfileirados. O Problema de Otimização de Pagamentos conta com uma busca por 2^n combinações, enquanto que, no modelo de Bech e Soramäki, a busca se limita a n combinações.

Nesse caso, observa-se $\lambda_A=2$, $\lambda_B=1$ e $\lambda_C=0$. Isso ocorre por que o acréscimo marginal de recursos ao banco A aumentará, em cascata, os recursos do banco B permitindo o aumento da liquidação em igual valor ao banco C. Haverá, então, aumento na liquidação de pagamentos duas vezes o tamanho do acréscimo de recursos ao banco A.

Essa observação permite o agrupamento dos bancos conforme o valor dos preços-sombra λ . Assim, conforme já observado, o preço-sombra relativo dos pagamentos μ_{ijk} / x_{ijk} dependerá exclusivamente do “grupo-origem” e do “grupo-destino” do pagamento. Todo pagamento proveniente de bancos com $\lambda=0$ e destinados a bancos com $\lambda=1$ terão $\mu_{ijk} / x_{ijk}=2$ pela igualdade em (3.12). Os pagamentos originários em $\lambda=0$ e destinados a $\lambda=1$ terão $\mu_{ijk} / x_{ijk}=0$ pela mesma relação. Ainda, todos os pagamentos que ocorrem dentro de um mesmo grupo têm $\mu_{ijk} / x_{ijk}=1$, independentemente de λ_i . Finalmente, observa-se que nenhum pagamento x_{ij} será feito quando a diferença $\lambda_j - \lambda_i < -1$, pois $\mu_{ij} > 0$. Assim, os pagamentos originários em banco do grupo $\lambda=2$ e destinados a $\lambda=0$ não serão realizados na solução ótima. Essa relação é intuitiva quando se observa a figura abaixo.

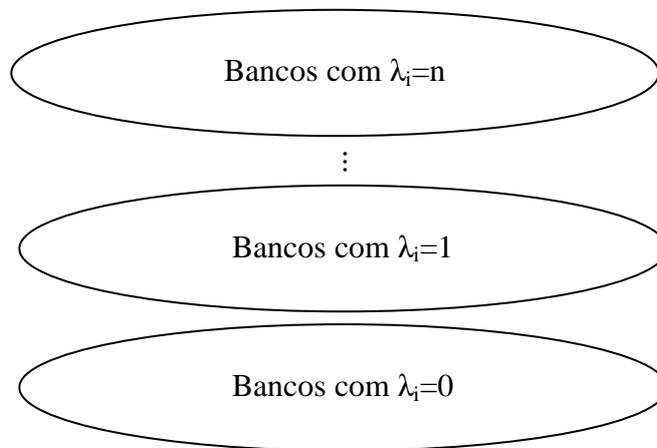


Figura 3.2

Todos os bancos pertencentes a determinado grupo possuem necessariamente pagamentos direcionados ao grupo imediatamente inferior e, caso também possuam pagamentos a bancos pertencentes a grupos mais de um nível abaixo, esses pagamentos terão menor prioridade (em função do próprio conceito do preço-sombra λ).

O agrupamento dos bancos em grupos conforme o valor de λ permite três observações importantes, que simplificam a busca de uma solução eficiente para o

Problema de Otimização de Pagamentos. Primeiro, na solução ótima **somente os pagamentos direcionados aos bancos pertencentes a grupos com menor valor de λ serão mantidos enfileirados**, ou seja, a liquidez existente na rede de pagamentos é suficiente para liquidá-los todos dentro dos grupos, assim como aqueles direcionados a grupos com maior valor λ . Essa observação tem demonstração trivial, pois dois bancos não poderiam pertencer ao mesmo grupo se um mantivesse pagamento pendente com o outro¹⁴. Exemplificando, considere dois bancos que pertençam ao grupo $\lambda=1$ e um banco que mantenha pagamentos pendentes de liquidação com o outro. Então o efeito do acréscimo marginal de recurso ao banco se estenderá ao outro, que, por pertencer ao grupo $\lambda=1$, usá-lo-á na liquidação de suas obrigações. Assim, o efeito marginal do incremento de recursos ao primeiro banco é maior que 1.

A segunda observação permite **priorizar a liquidação dos pagamentos em função da sua maximização**. Em problemas de programação linear com solução inteira – $v \in \{0,1\}$ –, o montante de recursos de um banco é necessário para ajustar a diferença entre os valores de um fluxo de pagamentos, permitindo a liquidação de valor significativamente maior entre dois bancos. A liquidação dos pagamentos permite que a liquidez percorra os bancos em uma rede, possibilitando novas liquidações. A priori, porém, se desconhece o efeito dessa transferência de recursos para a liquidação de novos pagamentos, ou seja, não é possível determinar qual será o valor dos mesmos, gerados a partir do aumento de recursos. Adota-se, então, a hipótese de que todos os bancos com pagamentos enfileirados possuem a mesma probabilidade de liquidar igual valor de pagamentos, dado o incremento de recursos. Os bancos com maior valor λ , contudo, multiplicam essa probabilidade pelo efeito cascata da liquidação de seus pagamentos. Assim, aqueles com maior relação μ_{ijk} / x_{ijk} devem ser priorizados, direcionando os recursos para os agrupamentos com maior valor λ .

Finalmente, ao considerar que a prioridade dos pagamentos seja atribuída pela relação μ_{ijk} / x_{ijk} , então **todos os pagamentos com origem em um grupo λ_1 e destino λ_2 (podendo $\lambda_1 = \lambda_2$) terão, a priori, a mesma prioridade**. Essa hipótese trás grande simplificação para solucionar o problema primal (3.8) e dual (3.9), pois permite o agrupamento dos pagamentos para o cômputo dos preços-sombra e a determinação do

¹⁴ Pode-se estender o mesmo raciocínio para pagamentos direcionados a grupos com λ de maior valor.

valor ótimo do problema. Assim, considere todos os pagamentos do banco i para o banco $j - x_{ijk} -$ como se fossem um único pagamento \bar{x}_{ij} , conforme apresentado abaixo:

$$(3.14) \quad \bar{x}_{ij} = \sum_{k=1}^{N^*} x_{ijk}$$

Essa simplificação reduz consideravelmente o esforço para se encontrar os preços-sombra dos pagamentos e não altera o resultado colocado no modelo anterior, se considerar que v_{ijk} pode assumir qualquer valor no intervalo $[0,1]$. Dessa forma, $\bar{v}_{ij}(t)$ será a função de liquidação média dos pagamentos, conforme abaixo:

$$(3.15) \quad \bar{x}_{ij}(t)' \cdot \bar{v}_{ij}(t) = \sum_{k=1}^{N^*} [x_{ijk}(t) v_{ijk}(t)] = \left(\sum_{k=1}^{N^*} x_{ijk}(t) \right) \bar{v}_{ij}(t) = \bar{x}_{ij}(t)' \cdot \bar{v}_{ij}(t)$$

Observa-se também que $\bar{v}_{ij}(t) \in [0,1]$.

A restrição do problema primal na notação matricial também sofre algumas simplificações. Considere a matriz \bar{Q}_i com dimensão $B \times (B-1)$ e $\bar{Q}_{i,l}$ a l -ésima linha da matriz. A regra de formação é definida como:

$$(3.16) \quad \bar{Q}_{i,l} = \begin{cases} 1 * \bar{x}_{ij}(t) & \text{se } i = l \\ -1 * \bar{x}_{ij}(t) & \text{se } j = l \\ 0 * \bar{x}_{ij}(t) & \text{se } j \neq l \end{cases}$$

Em um sistema interbancário com 4 bancos a matriz \bar{Q}_1 tem a forma:

$$\bar{Q}_1 = \begin{vmatrix} \bar{x}_{12} & \bar{x}_{13} & \bar{x}_{14} \\ -\bar{x}_{12} & 0 & 0 \\ 0 & -\bar{x}_{13} & 0 \\ 0 & 0 & -\bar{x}_{14} \end{vmatrix}$$

A matriz \bar{Q} , com dimensão $B \times B(B-1)$ é definida por:

$$(3.17) \quad \bar{Q} = [\bar{Q}_1 \mid \bar{Q}_2 \mid \dots \mid \bar{Q}_B]$$

O problema primal de otimização torna-se então:

$$(3.18) \text{ Max } x' \cdot v$$

$$\text{sa.: } \begin{vmatrix} Q \\ I \end{vmatrix} \cdot v \leq \begin{vmatrix} B \\ 1 \end{vmatrix}$$

$$v \geq 0$$

A solução do problema acima requer um custo computacional significativamente menor devido à redução do número de variáveis. Enquanto no problema inicial havia $B \times (B-1) \times N^*$ variáveis, no problema simplificado há $B \times (B-1)$ variáveis. A solução do problema acima é tratável computacionalmente e indicará quais transferências serão integralmente liquidadas ($\bar{v}_{ij} = 1$) e em quais transferências bilaterais serão parcialmente liquidadas ($\bar{v}_{ij} < 1$).

Devido à agregação dos pagamentos, é necessário um algoritmo que apresente uma regra e indique quais deles serão efetivamente liquidados no processo, ou seja, um algoritmo que permita incorporar a impossibilidade de liquidação fracionada ($v_{i,j}(t) \in \{0, 1\}$) e procure obter a solução mais próxima possível da solução ótima do problema de programação linear.

4. O Algoritmo de Liquidação

Na solução do problema de liquidação de pagamentos:

$$(4.1) \quad \text{Max } x' \cdot v$$

$$\text{s.a: } \begin{vmatrix} Q \\ I \end{vmatrix} \cdot v \leq \begin{vmatrix} B \\ 1 \end{vmatrix}$$

$$v \geq 0$$

O vetor x é exógeno ao problema e o resultado da maximização é o vetor da função liquidação v , cujos valores de cada elemento v_{ij} pertencem ao intervalo $[0,1]$. Para $v_{ij} = 0$, então nenhum pagamento do banco i para o banco j será liquidado e, quando $v_{ij} = 1$, todos os pagamentos enfileirados entre os dois bancos serão liquidados. Quando, porém, existe $v_{ij} \in (0,1)$, então a solução é insatisfatória, dado que, em geral, bancos centrais podem sofrer restrições de ordem legais ou técnicas e não podem permitir o fracionamento das mensagens.

Essa seção propõe um algoritmo para maximização dos pagamentos com solução inteira – sem fracionamento – observando os conceitos colocados na seção anterior. O algoritmo se baseia em sucessivas otimizações do fluxo de pagamentos entre as diversas combinações dos bancos agrupados em pares na rede de pagamentos, reduzindo um problema de maior complexidade – uma maximização com todos os pagamentos e B restrições – em sucessivos problemas de menor complexidade – maximização dos pagamentos entre dois bancos com uma ou duas restrições.

Num problema de otimização inteira, o saldo de recursos dos bancos é necessário para compensar a diferença entre o valor dos pagamentos que ingressam no banco e os que foram emitidos por este. Exemplificando, considere que o banco 1 emite 3 pagamentos para o banco 2 – $x_{12}=\{20,50,30\}$ – e deste recebe 1 pagamento – $x_{21}=\{15\}$. Se nenhum dos bancos tivesse recurso, então nenhum pagamento seria liquidado, mas se o banco 2 possuir pelo menos 15 unidades monetárias, então liquidará o pagamento transferindo seus recursos para o outro banco. Se, contudo, o banco 1 possuir 5 unidades monetárias, então liquidará o primeiro pagamento e, ao final do processo da otimização, o banco 2 concentrará o saldo de ambos os bancos antes da liquidação. Dois pontos se destacam nesse exemplo: a importância de recursos nos bancos para o processo de liquidação e a migração dos recursos ao final do processo. Assim, após a liquidação

bilateral o banco que obtiver mais recursos terá mais chances de liquidar novos pagamentos, enquanto o seu par, que com menos recursos, torna-se mais dependente do ingresso de recursos de outros bancos para novas liquidações¹⁵.

No processo de liquidação bilateral, a formação dos pares pode ocorrer de forma aleatória ou de acordo com uma sequência predeterminada. O maior risco da aleatoriedade é o empoçamento de liquidez, quando os recursos se deslocam a bancos que não possuem outros pagamentos enfileirados. Nessa situação, os recursos tornam-se indisponíveis ao sistema que, com menos recursos para os pagamentos remanescentes, perde eficácia nas liquidações subsequentes. Como resultado, observa-se maior tempo de processamento e menor número de liquidações na rede de pagamentos.

Para aumentar a eficiência do processo de liquidação bilateral deve-se priorizar os pagamentos com menos chances de empoçamento. A priorização dos pagamentos com maior preço-sombra (maior relação μ_{ijk} / x_{ijk}) evita o risco de empoçamento, visto que esses bancos são os que possuem maior necessidade de recursos para efetuarem novos pagamentos. Considere o exemplo abaixo:

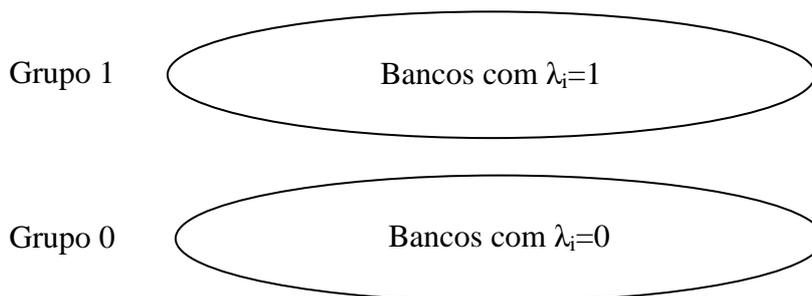


Figura 4.1

Nessa rede de pagamentos existem dois agrupamentos de bancos. Na seção anterior, observou-se que após a otimização dos pagamentos, os bancos do “balão” superior manterão pagamentos pendentes para os do balão inferior. Se a solução ótima quando $\nu \in \{0,1\}$ se aproximar da solução ótima do problema com $\nu \in [0,1]$, então haverá liquidação da maior parte dos pagamentos existentes dentro dos “balões” (pagamentos intragrupos) e dos pagamentos direcionados dos bancos do grupo 0 ($\lambda_i=0$) para os do grupo 1 ($\lambda_i=1$).

¹⁵ A liquidação de pagamentos sem uso de recursos ocorre quando há coincidência de valores no fluxo de pagamentos. A probabilidade dessa situação depende da distribuição dos pagamentos e reduz quando se aumenta a dispersão dos valores.

O algoritmo proposto nessa seção para solução do problema (4.1) tem por finalidade criar uma sequência de otimizações bilaterais que aproximem a solução do problema com $\nu \in \{0,1\}$ à solução ótima quando $\nu \in [0,1]$.

Antes de se iniciar a apresentação do algoritmo desenvolvido é conveniente apresentar um procedimento mais simples, introduzindo as dificuldades para chegar-se a uma solução. Considere uma rede de pagamentos com as características da figura 4.1 e um algoritmo que se inicia com a liquidação dos pagamentos direcionados ao grupo 1, ou seja, através da aplicação dos recursos dos bancos do grupo 0 nos pagamentos que possuem maior preço-sombra ($\mu_{ijk}/x_{ijk}=2$). Em seguida, os bancos do grupo 1 aplicam os recursos na liquidação dos pagamentos com seus pares dentro do grupo ($\mu_{ijk}/x_{ijk}=1$). Quando nenhum outro pagamento for liquidado no grupo 1, os recursos são utilizados para liquidar os pagamentos direcionados ao grupo 0 ($\mu_{ijk}/x_{ijk}=0$), que poderá utilizá-lo para liquidar os pagamentos dentro do grupo ($\mu_{ijk}/x_{ijk}=1$) e, posteriormente, reenviá-lo para o grupo 1 ($\mu_{ijk}/x_{ijk}=2$), repetindo o *loop* até a liquidação do maior número possível de pagamentos.

Este algoritmo induz a liquidação dos pagamentos conforme representado na figura abaixo, onde as setas largas representam a liquidação dos pagamentos e a seta fina indica a sequência das etapas.

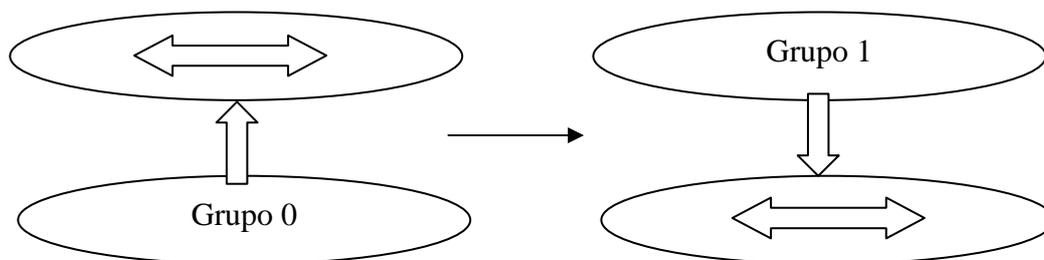


Figura 4.2

A sequência apresentada, contudo, é ineficiente. Ocorre que o fluxo dos pagamentos não é simetricamente distribuído entre os bancos de um mesmo grupo, ou seja, é possível (e muito provável) que existam bancos com posição multilateral negativa dentro dos agrupamentos. Esses bancos possuem maior necessidade de recursos que seus pares, que concentrarão a liquidez no grupo após sucessivas maximizações. Assim, embora a liquidez presente no sistema possibilite liquidar a maioria dos pagamentos dentro dos níveis, a simples transferência de recursos para o outro nível não garante que

as liquidações dos pagamentos ocorrerão de forma eficiente. Para melhorar o processo, é preciso que os recursos sejam direcionados para os bancos que possuem saldo multilateral negativo nos pagamentos dentro do grupo, possibilitando maior capacidade de liquidar seus pagamentos.

Outra fonte de ineficiência no algoritmo da figura 4.2 ocorre na liquidação dos pagamentos direcionados ao grupo 0 (parte direita da figura). Nessa etapa, os bancos cessionários são credores líquidos do fluxo de pagamentos e alguns bancos poderão não ter mais pagamentos pendentes com o outro grupo. Se esses bancos também tiverem saldo multilateral positivo, então poderão concentrar os recursos empochando a liquidez. Dessa forma, um algoritmo eficiente deve garantir que os recursos possam ser novamente transferidos ao outro grupo, direcionando-os preferivelmente aos bancos com saldo multilateral negativo.

A figura abaixo esboça a mudanças necessárias para corrigir as ineficiências apresentadas acima.

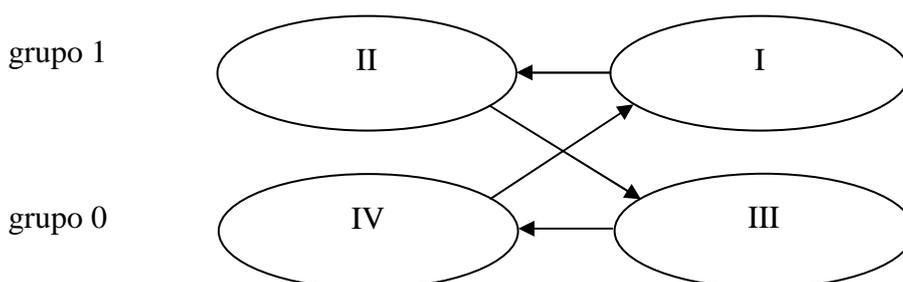


Figura 4.3

Para evitar prolixidade na descrição do novo algoritmo, todo o direcionamento dos pagamentos **de um grupo para outro** objetivando a migração dos recursos será denominado **transferência**. Com mesmo propósito, o direcionamento dos pagamentos para bancos **dentro de um mesmo grupo** será chamado **redistribuição**.

Na figura 4.3, cada grupo é dividido em dois conjuntos não necessariamente disjuntos de bancos. O conjunto I é constituído pelos bancos que possuem saldo multilateral negativo nos pagamentos dentro do grupo 1. O conjunto II é o seu complementar. O conjunto IV é constituído pelos bancos que possuem pagamentos direcionados ao grupo I e o conjunto III é formado por uma sequência de bancos ligada direta ou indiretamente ao conjunto IV (a formação do conjunto III será descrita adiante). Observa-se também que a liquidação dos pagamentos entre os grupos ocorre

em “X” e que os pagamentos se concentram nos conjuntos II e IV durante a etapa redistribuição.

A sequência apresentada acima corresponde ao caminho da liquidez na rede de pagamentos, iniciando no conjunto I e percorrendo todos os conjuntos até retornar ao conjunto de partida novamente. Assim a etapa preliminar ao *loop* é a transferência dos recursos no conjunto I. Observa-se também que o conjunto I tende a perder elementos para o conjunto II com a repetição da sequência acima, pois os pagamentos dentro do mesmo grupo se esgotam mais rapidamente que os pagamentos direcionados ao grupo 0. Quando os pagamentos se exaurem completamente, os conjuntos I e II se fundem em um único conjunto. Finalmente, o *loop* acima se esgota somente quando não houver mais pagamentos do grupo 0 direcionado ao grupo 1, ou seja, se os pagamentos direcionados ao conjunto I se exaurirem, então serão direcionados ao conjunto II¹⁶.

O detalhamento desse processo reside na descrição de três rotinas¹⁷: transferência, redistribuição e formação dos conjuntos III e IV. Através do uso sucessivo dessas rotinas é aplicada a sequência exibida na figura 4.2.

a. Transferência:

Essa etapa se propõe a transferir os recursos de um grupo a outro e há duas formas de promover a transferência através de maximizações. Na primeira, maximiza-se o **valor do fluxo de pagamentos** do banco i para o banco j, ou seja:

$$(4.2) \quad \text{Max} \sum_k x_{i,j,k} v_{i,j,k} - \sum_l x_{j,i,l} v_{j,i,l}$$

$$\text{s.a:} \quad \sum_k x_{i,j,k} v_{i,j,k} - \sum_l x_{j,i,l} v_{j,i,l} < S_i$$

Nesse caso, há necessidade de somente uma restrição de liquidez – para o banco i – pois se maximiza o ingresso de recursos no outro banco.

Distintamente, pode-se optar por maximizar **valor dos pagamentos** do banco i para o banco j, ou seja:

$$(4.3) \quad \text{Max} \sum_k x_{i,j,k} v_{i,j,k}$$

¹⁶ Esse cenário pressupõe que o conjunto I ainda exista.

¹⁷ Essas são as rotinas após a solução do problema pelo PPL para a obtenção dos preços-sombras λ_i e $\mu_{i,j}$ pela solução do problema em .

$$\begin{aligned} & \sum_k x_{i,j,k} v_{i,j,k} - \sum_l x_{j,i,l} v_{j,i,l} \leq S_i \\ \text{s.a:} & - \sum_k x_{i,j,k} v_{i,j,k} + \sum_l x_{j,i,l} v_{j,i,l} \leq S_j \end{aligned}$$

No segundo caso, é necessário incluir uma segunda restrição para o outro banco. Para entender melhor a diferença entre as duas alternativas, considere como exemplo dois bancos {1,2} sendo $S_1=10$; $x_{12}=20$; $x_{21}=40$. Caso opte por (4.2), então não haverá transferência pois o saldo do banco 1 é insuficiente para liquidar o pagamento e, sendo $x_{12} < x_{21}$, a compensação entre os pagamentos torna negativo o fluxo direcionado ao banco 2. Ao contrário, se o problema for colocado da mesma maneira que em (4.3), haverá a transferência se $S_2 \geq 20$. Contudo, no segundo caso os recursos migrarão no sentido inverso ao desejado.

Ambas as alternativas foram avaliadas e **optou-se por modelar a rotina transferência como em (4.3)**. Nos testes efetuados, verificou-se que em ambos os casos os recursos são direcionados para o outro grupo¹⁸, porém a liquidação dos pagamentos é substancialmente maior quando o problema é modelado da mesma maneira que em (4.3).

b. Redistribuição:

A rotina redistribuição inicia-se quando os recursos estão concentrados em alguns bancos e almeja-se maximizar a **liquidação do maior valor em pagamentos** entre dois bancos do mesmo nível, ou seja:

$$\begin{aligned} (4.4) \quad & \text{Max} \sum_k x_{i,j,k} v_{i,j,k} + \sum_l x_{j,i,l} v_{j,i,l} \\ \text{s.a:} & \sum_k x_{i,j,k} v_{i,j,k} - \sum_l x_{j,i,l} v_{j,i,l} < S_i \\ & - \sum_k x_{i,j,k} v_{i,j,k} + \sum_l x_{j,i,l} v_{j,i,l} < S_j \end{aligned}$$

Essa modelagem requer duas restrições de liquidez, evitando que ambos os bancos apresentem saldo negativo no processo.

c. Formação dos conjuntos III e IV:

¹⁸ Embora tenha sido observado que pontualmente os recursos possam caminhar no sentido contrário ao desejado.

O conjunto IV é formado pelos bancos que mantêm pagamentos enfileirados aos bancos do outro grupo. A seleção tem preferência pelos bancos que mantenham saldo multilateral positivo com seus pares, permitindo maior número de liquidações de pagamentos intragrupo, e que tenham pagamentos direcionados ao conjunto I.

Após a seleção dos bancos do conjunto IV, a formação do conjunto III inicia-se pelos bancos que possuem pagamentos pendentes àqueles e prosegue com a inclusão de outros bancos do nível com pagamentos direcionados aos bancos já incluídos no conjunto III. Esta sequência é repetida enquanto houver bancos no agrupamento com pagamentos direcionados aos demais bancos incluídos no conjunto. Exemplificando o procedimento, considere uma rede de pagamentos como a da figura 4.3. No grupo 0 há 5 bancos (A, B, C, D e E) e somente o banco A mantém pagamentos enfileirados para o conjunto I. Dessa forma, o banco A compõe o conjunto IV e recebe pagamentos dos bancos B e C, que são então incluídos no conjunto III. Esses bancos também recebem pagamentos dos bancos A e D, que não recebem pagamentos do banco E. O conjunto III é formado, então, por {A, B, C, D}.

Analisando o exemplo acima, observa-se que, ao contrário dos conjuntos I e II, os conjuntos III e IV não são disjuntos – pois o banco A pertence a ambos – e sua união está contida no grupo, mas o contrário não é verdadeiro – o banco E pertence ao grupo, mas não aos conjuntos.

O procedimento descrito acima permite identificar bancos que representam risco de empoçamento. Os bancos do grupo que não são incluídos nos conjuntos são denominados **desconectados** e retêm os recursos dentro do mesmo. Assim, é preciso postergar a liquidação dos pagamentos direcionados a esses bancos.

Por fim, observa-se que há uma hierarquia ou “*rank*” na entrada dos bancos no conjunto III. Os primeiros bancos (B e C) são os que possuem pagamentos direcionados diretamente ao banco A e possuem *rank* igual a 1. Em seguida, entram os bancos A e D com *rank* igual a 2. A tabela abaixo resume a classificação dos bancos no exemplo, colocando *rank* igual a 0 para o banco do conjunto IV.

<i>Conjunto</i>	<i>Bancos</i>	<i>Rank</i>
IV	A	0
III	B e C	1
III	A e D	2

Tabela 4.1

A lógica na formação do conjunto III é criar um “caminho” para que os recursos sejam utilizados nas liquidações dos pagamentos intragrupo, mas garantindo que se concentrarão nos bancos do conjunto IV e não se desviarão para os bancos desconectados. No exemplo, os bancos do conjunto II poderão transferir os recursos (rotina transferência) para os bancos A e D. Esses bancos podem usar os recursos para liquidar pagamentos entre si (rotina redistribuição) e, em seguida, os transferirem para os bancos B e C que, por sua vez, irá usá-los na liquidação de seus pagamentos, sendo primeiro entre si e depois para o banco A.

O sucesso para garantir a etapa seguinte – transferência dos recursos para o outro grupo – exige a adoção de algumas precauções adicionais. Se houver mais de um banco no conjunto IV, é possível que os recursos se concentrem naqueles cujos pagamentos ao outro grupo somem valor insignificante frente ao valor dos recursos existentes na rede de pagamentos. Nessa situação, parte expressiva dos recursos não será enviada ao outro grupo, reduzindo a eficiência das etapas subsequentes. O saldo final de cada banco deve, então, ser limitado à sua capacidade de liquidar pagamentos ao outro grupo¹⁹.

De forma análoga, o mesmo raciocínio pode ser estendido aos bancos do conjunto III, evitando que estoquem recursos que não conseguirão transferir aos bancos de *rank* imediatamente inferior. Os limites de recursos destinados aos bancos dos conjuntos III e IV são obtidos através da sequência abaixo:

1. Imposição de limites aos bancos do conjunto IV em função do valor dos pagamentos direcionados aos bancos do conjunto I;
2. Ordenamento descendente dos **bancos do conjunto IV** pelo valor dos limites do item anterior;

¹⁹ A exceção a essa regra seria a situação onde a soma dos pagamentos destinados ao outro grupo for menor que a liquidez existente. Nesse caso, **cada banco** deverá adquirir pelo menos o valor igual à soma dos pagamentos direcionados ao outro grupo.

3. Ordenamento descendente dos **bancos com rank igual a 1** pelo valor da soma dos pagamentos provenientes do conjunto II e direcionados a esses bancos;
4. Para o **primeiro banco do ordenamento acima**, obtenção do mínimo entre a soma dos pagamentos direcionados ao primeiro banco do *rank* imediatamente inferior e o limite daquele banco. O valor obtido compõe o limite do banco do conjunto III e subtrai o limite do banco do *rank* inferior²⁰, ajustando-o para a determinação do limite dos outros bancos (etapa 6).
5. Para o mesmo banco com *rank* 1, repetição da etapa 4 até o esgotamento dos bancos do conjunto IV;
6. Para os demais bancos do *rank* 1, repetição das etapas 4 e 5, usando-se o limite ajustado na etapa 4 para o banco cessionário.
7. Repetição das etapas 3 a 6, substituindo os bancos pelos seus pares com *rank* imediatamente superior.

Ao término da etapa 7, todos os bancos possuirão limites definidos.

Para exemplificar a sequência acima, considere ainda a tabela 4.1 e a figura 4.3 e assumo $\sum_j \sum_k x_{A,j,k} = 100$, $j \in$ conjunto I. Considerando que o banco B receba mais pagamentos do conjunto II que o banco C (etapa 3) e que $\sum_j \sum_k x_{B,A,k} = 40$ e $\sum_j \sum_k x_{C,A,k} = 120$.

Pelas etapas 4 e 5, obtém-se:

Limite de B: $\min(100;40) = 40$

Limite de C: $\min(60;120) = 60$

Para os bancos do *rank* 2, considere que o banco A receba mais pagamentos do conjunto II que o banco D e que:

$$\sum_j \sum_k x_{A,B,k} = 50; \sum_j \sum_k x_{A,C,k} = 30; \sum_j \sum_k x_{D,B,k} = 40 \text{ e } \sum_j \sum_k x_{B,A,k} = 80$$

O limite de A (no *rank* 2) é: $\min(40;50) + \min(60;30) = 70$;

²⁰ Nessa etapa, o limite do banco do *rank* inferior é ajustado considerando que haverá a transferência dos recursos. Assim, ajusta-se o limite (saldo residual) para outras transferências dos demais bancos do conjunto III.

O limite de D é: $\min(0;40) + \min(30;80) = 30$;

A tabela 4.2 abaixo resume os resultados obtidos pelo exemplo.

<i>Conjunto</i>	<i>Bancos</i>	<i>Rank</i>	<i>Limite</i>
IV	A	0	100
III	B	1	40
III	C	1	60
III	A	2	70
III	D	2	30

Tabela 4.2

No algoritmo, os bancos do conjunto II deverão direcionar a liquidação dos pagamentos, preferencialmente, para os bancos de maior *rank*. No exemplo acima, as liquidações são direcionadas para os bancos A e D.

A **rotina transferência** deve considerar os limites dos bancos na liquidação dos pagamentos. A adaptação é facilmente introduzida no modelo (4.3) ajustando-se as restrições do problema aos limites dos bancos, conforme apresentado abaixo para a transferência genérica do banco i para o banco j:

$$(4.5) \quad \text{Max} \sum_k x_{i,j,k} v_{i,j,k}$$

$$\sum_k x_{i,j,k} v_{i,j,k} - \sum_l x_{j,i,l} v_{j,i,l} \leq \min(S_i, \max(0, \text{Limite bco } i - S_i))$$

s.a:

$$-\sum_k x_{i,j,k} v_{i,j,k} + \sum_l x_{j,i,l} v_{j,i,l} \leq \min(S_j, \max(0, \text{Limite bco } j - S_j))$$

De forma análoga, as liquidações entre bancos do mesmo *rank* ocorrem através da rotina redistribuição, conforme apresentado abaixo:

$$(4.6) \quad \text{Max} \sum_k x_{i,j,k} v_{i,j,k} + \sum_l x_{j,i,l} v_{j,i,l}$$

$$\sum_k x_{i,j,k} v_{i,j,k} - \sum_l x_{j,i,l} v_{j,i,l} \leq \min(S_i, \max(0, \text{Limite bco } i - S_i))$$

s.a:

$$-\sum_k x_{i,j,k} v_{i,j,k} + \sum_l x_{j,i,l} v_{j,i,l} \leq \min(S_j, \max(0, \text{Limite bco } j - S_j))$$

Por fim, a rotina de formação dos conjuntos III e IV apresentada está intrinsecamente associada à rede de pagamentos com dois grupos de preço-sombra λ

($\lambda=0$ e $\lambda=1$). A generalização da rotina desenvolvida para redes com $n+1$ grupos foi feita através do agrupamento dos bancos em dois grupos, sendo $\lambda=n$ e $\lambda=\{0,1,\dots,n-1\}$, conforme a figura abaixo:

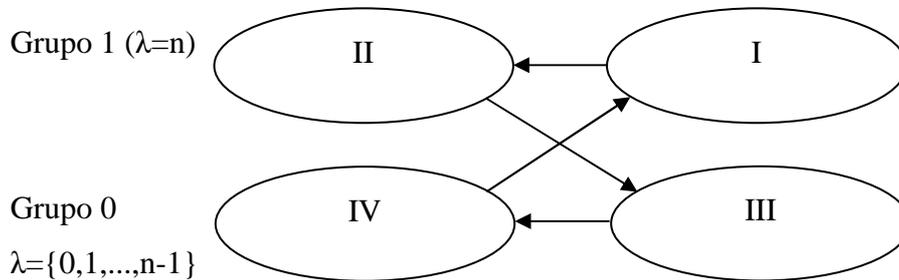


Figura 4.4

Todas as etapas se mantêm conforme apresentadas no corpo da seção, sendo que os conjuntos III e IV incluirão os bancos com todos os valores de λ , à exceção de $\lambda=n$. Quando não houver mais pagamentos direcionados aos bancos deste grupo 1, então os bancos do grupo $\lambda=n$ são substituídos pelos bancos com $\lambda=n-1$, que passam a compor o grupo 1. Esse processo se repete até não haver mais pagamentos do grupo IV para o grupo 1 composto pelos bancos com $\lambda=1$.

Ao final do *loop* apresentado na figura 4.4, haverá ainda pagamentos que poderão ser feitos a alguns bancos que foram “desconectados” do processo. Para liquidar esses pagamentos utiliza-se a rotina redistribuição considerando todos pagamentos ainda pendentes de liquidação²¹.

²¹Ocorre que, no processo de liquidação, alguns pagamentos não são executados pelo saldo insuficiente de recursos entre as partes. A inclusão de todos os pagamentos proporciona uma segunda chance para que pagamentos não realizados durante o loop sejam liquidados.

5. Performance do algoritmo:

Nesta seção, avalia-se a performance do modelo desenvolvido com a aplicação do algoritmo apresentado na seção anterior. Dada as características da solução, esse modelo será denominado Bilateral Assistido, e os resultados obtidos serão comparados com o com o valor ótimo da solução do problema por programação linear (PPL) – modelado em (3.9) – e com outros modelos: LBTR, Bech-Soramäki e Bilateral Aleatório.

No Sistema LBTR os pagamentos são liquidados pelos valores brutos e cada banco está sujeito à restrição de liquidez. A liquidação dos pagamentos, contudo, pode ou não estar condicionada à restrição sequencial *FIFO*, onde todos pagamentos permanecem em fila aguardando a liquidação do primeiro. Nessa seção, o ordenamento dos pagamentos não pertence à restrição de liquidação nesse sistema, sendo os recursos direcionados ao próximo pagamento enfileirado quando não há liquidez suficiente para o primeiro. A razão dessa escolha está relacionada à condição de indiferença à prioridade dos pagamentos pelos bancos para o desenvolvimento desse estudo e pelo fato da ausência da restrição sequencial proporcionar maior eficiência no emprego da liquidez existente na rede de pagamentos. O fluxograma desse modelo LBTR encontra-se apresentado no apêndice 1.

O modelo denominado Bech-Soramäki foi apresentado pelos autores em 2001 para a liquidação dos pagamentos que obedecem ao ordenamento *FIFO* e permite a compensação multilateral entre os mesmos. Em caso de resultado multilateral negativo com insuficiência de recursos de algum banco, retira-se o último pagamento da fila de algum banco inadimplente, recalculando o resultado multilateral. Quando todos os bancos tiverem recursos para liquidar suas posições, ocorre a liquidação simultânea dos pagamentos. O fluxograma desse modelo encontra-se no apêndice 2.

O terceiro modelo, denominado Bilateral Aleatório, é semelhante ao modelo Bilateral Assistido e a principal diferença ocorre na escolha aleatória dos pares de bancos nesse modelo. A comparação entre esses modelos de liquidação bilateral com compensação avalia diretamente os benefícios introduzidos na seção 4 e, conseqüentemente, os avanços gerados pela incorporação dos conceitos apresentados na seção 3 no processo de liquidação bilateral dos pagamentos.

A soma das liquidações dos pagamentos em cada modelo foi comparada com o resultado da solução ótima do PPL²², conforme a fórmula apresentada abaixo, e usada para avaliar a performance dos modelos.

$$(5.1) \text{ Performance} = \frac{\text{Valor dos pagamentos liquidados}}{\text{Valor ótimo do PPL}}$$

A elaboração e distribuição dos pagamentos ocorreu de forma aleatória, parametrizando-se o número de bancos na rede de pagamentos (N), o número máximo de pagamentos que um banco pode fazer a outro (P), o valor máximo que um pagamento pode assumir (VMax) e o saldo inicial dos recursos de cada banco (S_i). Controla-se, assim, o processo de geração dos dados, permitindo observar a performance do algoritmo vis-à-vis variações:

- a. na distribuição dos pagamentos;
- b. na distribuição dos recursos (alocação dos saldos dos bancos);
- c. na amplitude dos valores máximo e mínimo dos pagamentos.

Nos itens abaixo serão descritos os detalhamentos para a elaboração dos cenários em que os algoritmos foram avaliados, assim como os resultados obtidos.

a. Variações na distribuição dos pagamentos:

Nesse exercício, avalia-se a performance do algoritmo em redes de pagamentos com diferentes níveis de simetria. Redes com alta simetria de pagamentos, ou seja, onde todos os bancos recebem e transferem valores de pagamentos da mesma ordem de grandeza possuem alto poder de compensação entre pagamentos e recebimentos. As redes de pagamentos mais assimétricas, ao contrário, concentram os pagamentos em poucos bancos e possuem menor poder de compensação. Objetivando avaliar a performance do modelo, foram criadas três regras de formação de pagamentos.

Na primeira regra – **regra 1** – todos os N bancos distribuem P pagamentos a cada par. O valor de cada pagamento é uma variável aleatória com distribuição uniforme entre [1,VMax].

Exemplificando, considere uma rede de pagamentos com três bancos {1, 2, 3}, onde cada instituição efetua 5 pagamentos, cujo valor é limitado a \$10, para cada par. Os

²² A comparação do resultado com o problema de programação linear e não com o ótimo do problema impondo $v=\{0,1\}$, deve-se ao alto tempo de processamento da alternativa.

parâmetros desse problema são: $N=3$, $P=5$ e $V_{Max}=10$. Uma possível rede de pagamentos é ilustrada na tabela abaixo:

	1	2	3
1	-	{3;8;10;6;2}	{1;3;2;9;1}
2	{3;5;1;3;8}	-	{9;4;5;1;2}
3	{8;1;6;4;9}	{4;7;2;1;3}	-

Tabela 5.1

Na tabela, considere que os bancos nas linhas efetuem os pagamentos aos seus pares nas colunas. Assim, o banco 1 recebe \$48 e transfere \$45, com posição multilateral \$3. Similarmente, os bancos 2 e 3 têm posição multilateral \$5 e -\$8. O saldo inicial dos pagamentos é o vetor $S=\{4,5,2\}$, indicando que o banco 3 não teria recursos para liquidar sua posição multilateral líquida.

Na segunda regra de formação dos pagamentos – **regra 2** – aumenta-se a assimetria na distribuição dos pagamentos, atribuindo-se as seguintes probabilidades à formação dos pagamentos entre dois bancos:

- 30% - não haverá pagamentos ;
- 40% - haverá $(P/5)$ pagamentos²³;
- 30% - haverá P pagamentos.

Na última regra de formação – **regra 3** –, considere V uma variável aleatória com distribuição uniforme no intervalo $[1,P]$. Os pagamentos entre cada par de bancos são constituídos com a seguinte probabilidade:

- 60% - não haverá pagamentos ;
- 30% - haverá $V/5$ pagamentos²⁴;
- 10% - haverá V pagamentos

Em ambas as regras, o valor de cada pagamento é uma variável aleatória uniforme no intervalo $[1,V_{Max}]$.

A definição dos valores dos saldos dos bancos no início do processo (S_i) é uma variável aleatória com distribuição uniforme entre $[1,V_{Max}]$ em todas as regras apresentadas.

²³ Haverá arredondamento para número inteiro quando o número de pagamentos não for múltiplo de 5.

²⁴ Idem comentário 24.

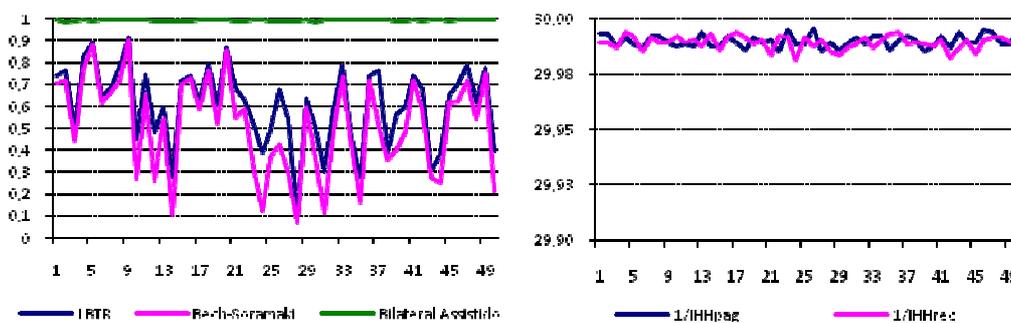
O índice de Herfindahl-Hirschman (IHH) foi utilizado para monitoramento do grau de simetria da rede de pagamentos, sendo usado para monitorar a concentração dos pagamentos (IHHpag) e dos recebimentos (IHHrec). A fórmula é apresentada abaixo, sendo o banco cedente, o banco cessionário e os pagamentos representados pelos índices i, j e k, respectivamente.

$$(5.2) \quad \text{IHHpag} = \sum_i \left(\frac{\sum_j \sum_k x_{i,j,k}}{\sum_i \sum_j \sum_k x_{i,j,k}} \right)^2 ;$$

$$(5.3) \quad \text{IHHrec} = \sum_j \left(\frac{\sum_i \sum_k x_{i,j,k}}{\sum_i \sum_j \sum_k x_{i,j,k}} \right)^2$$

Um valor alto do IHHpag, por exemplo, indica que os pagamentos têm origem em poucos bancos. Valores baixos do IHHrec sugerem que os pagamentos estão direcionados aos diversos bancos da rede²⁵. Quando ambos os índices têm valores baixos, há maior poder de compensação. Quando, ao contrário, ambos têm valores altos, a probabilidade de compensação no fluxo de pagamentos se reduz.

Os gráficos abaixo apresentam os resultados obtidos em 50 ensaios para a regra 1 com os parâmetros: N=30; P=30; VMax=100.



Gráficos 5.1 e 5.2

O gráfico à esquerda apresenta o percentual de liquidação dos pagamentos comparado ao valor ótimo da solução do PPL e o gráfico a direita exhibe os valores do inverso dos índices IHH para os pagamentos e recebimentos. Neste gráfico observa-se

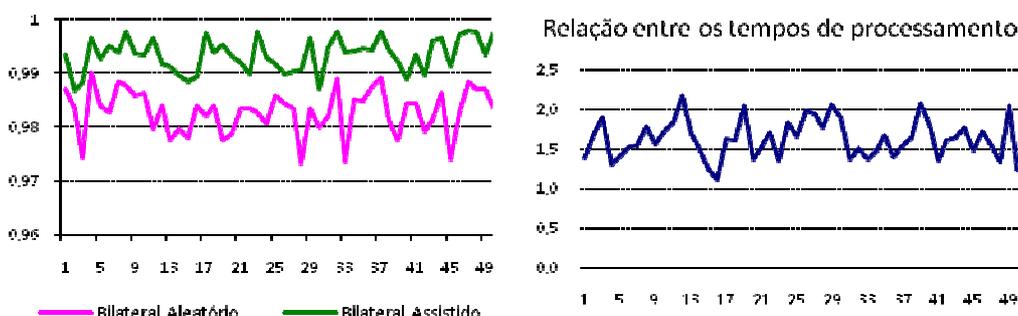
²⁵ O valor do IHH pertence ao intervalo $[1, 1/N]$, sendo 1 o cenário de maior assimetria e $1/N$ o de maior simetria.

que ambos os valores são muito próximos ao limite máximo (o limite superior é 30), indicando alta simetria na distribuição do fluxo de recursos. Nesse cenário, o modelo Bilateral Assistido conseguiu a liquidação da quase totalidade dos pagamentos. As soluções pelos modelos LBTR e Bech-Soramäki apresentaram resultados inferiores e grande volatilidade, conforme observado na tabela abaixo:

	<i>Média</i>	<i>Desvio-Padrão</i>
LBTR	60,5%	18,1%
Bech-Soramäki	51,5%	22,1%
Bilateral Assistido	99,3%	0,3%

Tabela 5.2

A comparação com o modelo Bilateral Aleatório é apresentada no gráfico abaixo.

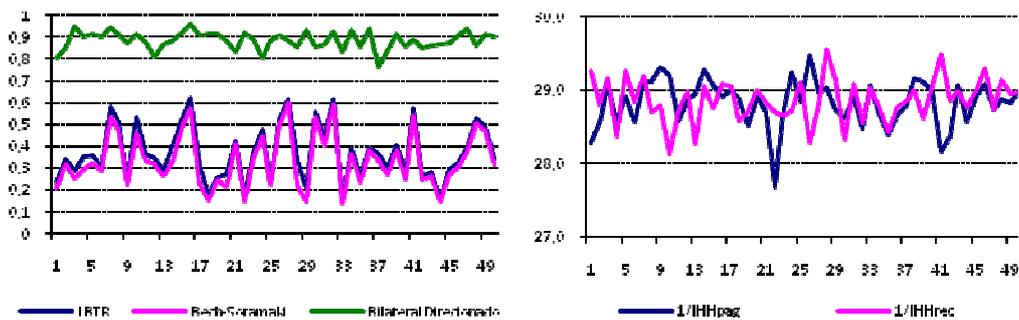


Gráficos 5.3 e 5.4

O gráfico 5.3 revela que ambos os modelos apresentaram alto rendimento, sendo que o modelo Assistido foi ligeiramente superior ao Aleatório em todos os experimentos realizados. Esse resultado indica a ausência de problemas de empocamento em redes com grande homogeneidade entre pagamentos e recebimentos.

No gráfico 5.4, observa-se a relação entre os tempos de processamento de ambos os modelos, sendo o Bilateral Aleatório 1,63 vezes mais demorado.

Os resultados para os dados gerado pela regra 2 são apresentados nos gráficos abaixo:



Gráficos 5.5 e 5.6

O gráfico 5.5 mostra a deterioração da performance de todos os modelos nesse novo cenário. Nos modelos LBTR e Bech-Soramäki houve, porém, redução da volatilidade, que aumentou no Bilateral Direcionado. A principal razão da redução da performance do algoritmo é a menor simetria entre pagamentos e recebimentos, observada no gráfico à direita. As médias do inverso dos índices IHHpag e IHHrec foram respectivamente 28,83 e 28,86 (contra 29,99 na regra anterior).

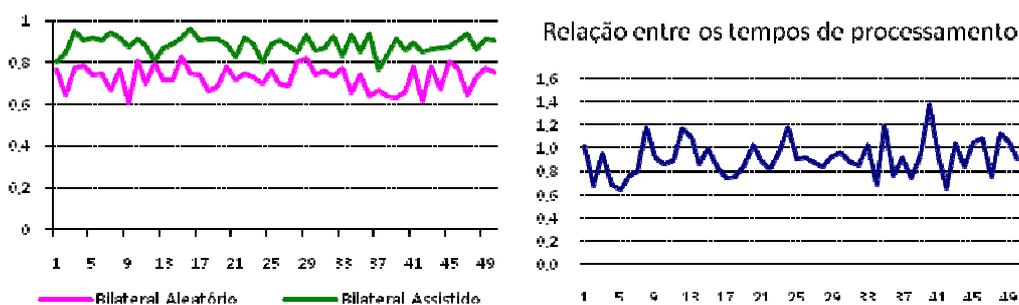
A assimetria reduz a compensação entre pagamentos e recebimentos (afetando o modelo Bilateral Assistido) e, ao direcionar os pagamentos de forma assimétrica, concentra a liquidez do sistema em bancos que se “desconectam” da rede de pagamentos, reduzindo a performance dos modelos LBTR e Bech-Soramäki.

A tabela 5.2 apresenta a média e desvio-padrão dos modelos do gráfico.

	<i>Média</i>	<i>Desvio-Padrão</i>
LBTR	36,7%	12,9%
Bech-Soramäki	33,6%	12,8%
Bilateral Direcionado	88,5%	4,2%

Tabela 5.2

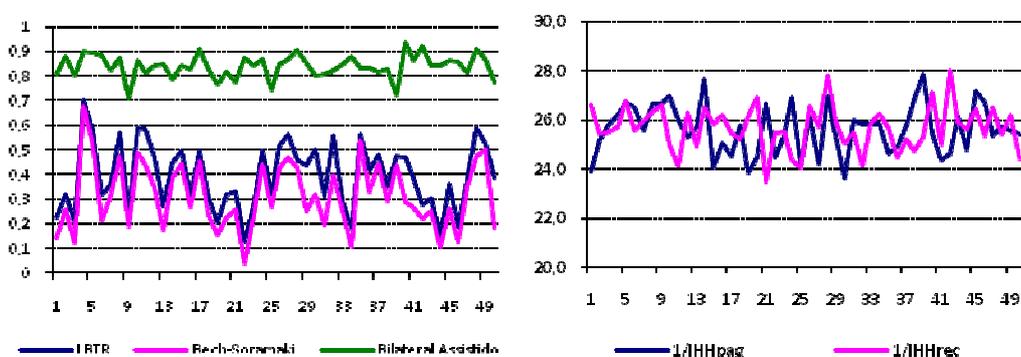
A comparação com o modelo Bilateral Aleatório é apresentada nos gráficos a seguir:



Gráficos 5.7 e 5.8

Pelo gráfico 5.7, observa-se que a diferença entre a performance dos modelos aumentou. Enquanto no modelo Bilateral Assistido a liquidação dos pagamentos caiu para 88,5% do valor ótimo (queda de 10,62% em relação a média anterior), no modelo Bilateral Aleatório a média caiu para 72,6% (redução de 26,1%). O gráfico 5.8 revela que a interrupção precoce do processo de otimização, decorrente do empocamento da liquidez, reduziu a relação entre os tempos de processamento dos modelos, tendo a rotina de otimização do Bilateral Aleatório consumido 91,6% do processamento do outro modelo.

Finalmente, os gráficos abaixo apresentam a comparação entre os modelos na terceira regra de formação dos pagamentos.

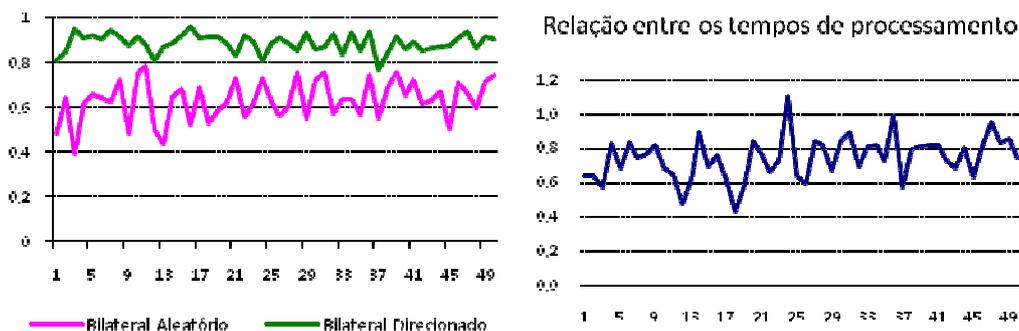


Gráficos 5.9 e 5.10

Nos gráficos verifica-se nova queda da performance do modelo Bilateral Assistido, cuja média caiu para 83,9% (queda de 5,3% em relação à regra 2). Os modelos LBTR e Bech-Soramäki não apresentaram uma variação significativa, tendo a média do primeiro aumentado para 39,2% e a do segundo caído para 31,3%.

O gráfico 5.10 evidencia a maior concentração de pagamentos e recebimentos na regra 3, sendo as médias 25,6 e 25,7 respectivamente. Contudo, a redução discreta no resultado do modelo mostra que o algoritmo foi eficiente quanto a evitar o direcionamento da liquidez para os bancos desconectados.

A comparação com o modelo Bilateral aleatório é apresentada abaixo.



Gráficos 5.11 e 5.12

A média da performance do modelo Bilateral Aleatório cai para 63,0% (redução de 13,2% em relação à performance anterior), evidenciando a forte correlação entre o aumento da assimetria do fluxo de pagamentos e a eficiência desse modelo. O tempo de processamento, contudo, cai para 74,7% do processamento do modelo Bilateral Assistido.

Em resumo, observa-se que o aumento da assimetria afeta todos os processos de liquidação de pagamentos. Nos processo de otimização bilateral de pagamentos, contudo, esses efeitos afetaram menos o algoritmo apresentado na seção 4 do que o modelo Bilateral Aleatório, apresentando melhor performance em todas as regras de formação de pagamentos. Quando a diferença na liquidação dos pagamentos foi irrelevante (regra 1), o modelo se destacou pelo menor tempo de processamento.

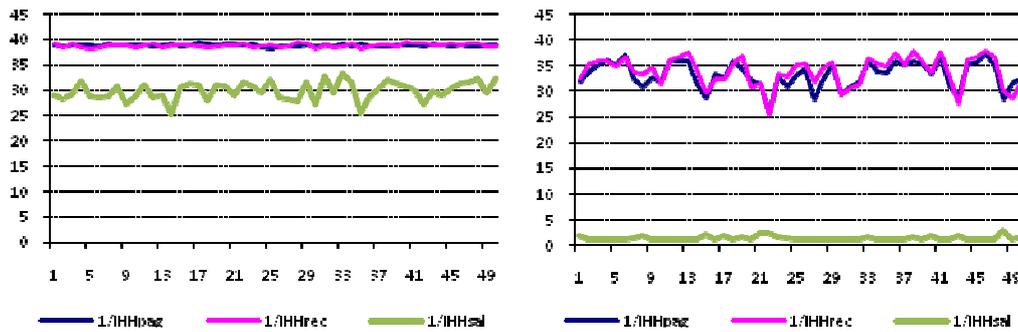
b. Variações na distribuição dos recursos:

A avaliação do efeito da distribuição dos saldos dos bancos no modelo visa observar a performance do algoritmo quando os recursos na rede de pagamentos estão concentrados em bancos que possuem poucos (ou nenhum) pagamentos com seus pares, havendo, conseqüentemente, menos recursos para liquidar os pagamentos. Esse cenário foi usando a regra 2 de formação de pagamentos (com os parâmetros $N=40$, $P=40$ e $V_{Max}=100$), seguido da liquidação dos pagamentos pelo modelo LBTR. Os modelos Bilateral Aleatório, Bilateral Assistido e Bech-Soramäki foram executados após o LBTR, sendo monitorada a capacidade de liquidar novos pagamentos, ou seja, a performance complementar desses modelos.

A distribuição dos saldos em recursos para pagamentos é avaliada pelo índice IHHsal, cuja fórmula é apresentada a seguir:

$$(5.3) \quad \text{IHHSaldo} = \sum_i \left(\frac{S_i}{\sum_i S_i} \right)^2;$$

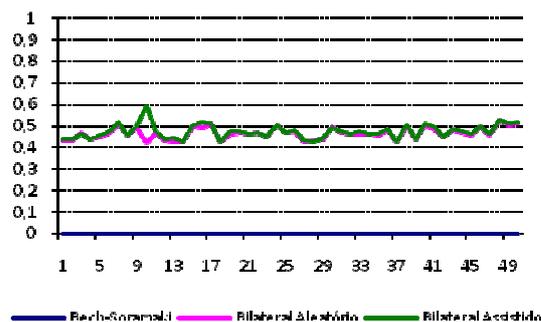
Os gráficos 5.13 e 5.14 exibem a rede de pagamentos antes e após a liquidação dos pagamentos pelo modelo LBTR.



Gráficos 5.13 e 5.14

A comparação pelos gráficos mostra claramente que o principal efeito do modelo LBTR é a concentração dos saldos em um pequeno número de bancos, cuja média do inverso do índice IHHsal caiu de 29,8 para 1,4²⁶. A distribuição do fluxo de pagamentos também torna-se mais assimétrica.

O gráfico abaixo mostra a performance complementar dos modelos na liquidação dos pagamentos remanescentes.



Gráficos 5.15

O modelo de Bech-Soramäki apresentou resultado próximo a zero devido à baixa quantidade de recursos nos bancos com pagamentos pendentes. Assim, a liquidação

²⁶ Nesse momento, observa-se que o inverso do índice IHH revela o número de entes equivalentes no sistema. O valor IHHsal menor que 2 revela que a liquidez se concentra de forma que um banco possua mais recursos que a soma dos recursos dos demais.

pelo algoritmo fica condicionada à coincidência da soma dos valores dos pagamentos, sujeitos ao ordenamento sequencial, no saldo multilateral para gerar compensação. A média desse modelo foi 0,01%, sendo a maior performance 0.17%.

Os modelos Bilateral Aleatório e Bilateral Direcionado mantiveram resultados muito próximos, também em função da alta concentração dos recursos. A liquidação dos pagamentos ocorre pela compensação entre pagamentos e recebimentos entre os pares de bancos, porém com vantagem sobre o modelo de Bech-Soramäki, por este estar condicionado ao ordenamento sequencial. Ambos os modelos apresentaram resultados abaixo dos observados em cenários anteriores, o que fica evidenciado na tabela abaixo:

	<i>Média</i>	<i>Desvio-Padrão</i>
Bech-Soramäki	0,01%	0,03%
Bilateral Aleatório	46,6%	2,8%
Bilateral Direcionado	47,3%	3,3%

Tabela 5.3

A relação entre os tempos de processamento entre ambos os modelos é apresentada no gráfico seguinte:



Gráfico 5.16

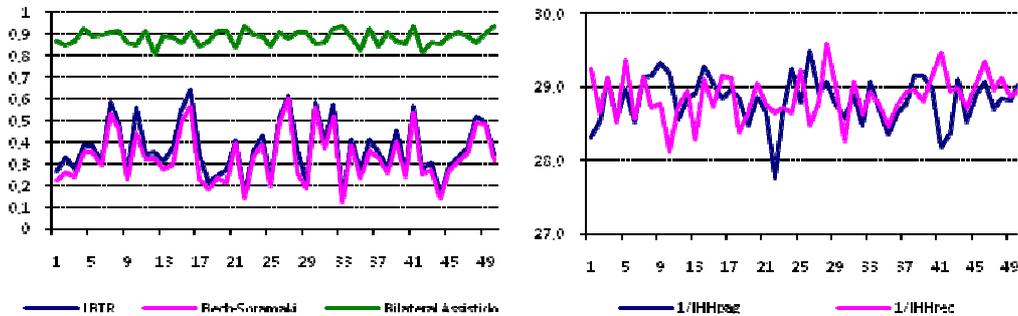
A média da relação entre o tempo de processamento foi 1,00 e desvio-padrão 0,26, indicando que ambos os modelos são igualmente satisfatórios em cenários de baixa liquidez entre os participantes.

c. Variações na amplitude entre os valores máximo e mínimo dos pagamentos:

O objetivo desse exercício é avaliar a performance dos modelos quando se aumenta a dispersão dos valores dos pagamentos. A regra 2 de formação de pagamentos foi

novamente utilizada, com os parâmetros $N=30$; $P=30$ e $V_{Max}=1000$ (a amplitude para a geração de pagamentos aumentou 10 vezes em relação ao exercício do item a).

A performance do algoritmo e a distribuição dos pagamentos é observada nos gráficos abaixo.



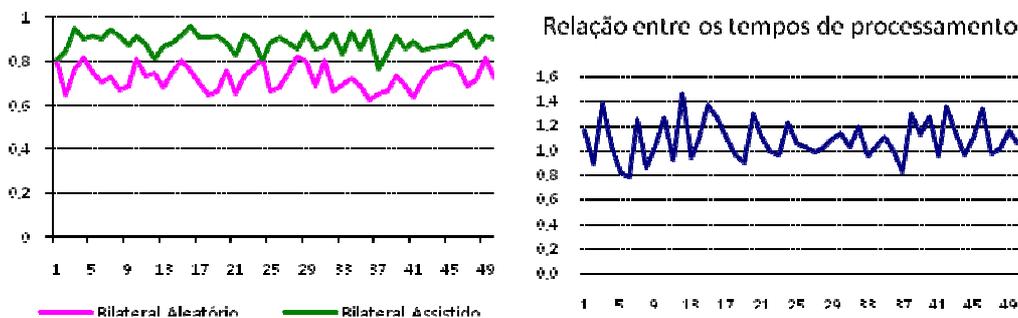
Gráficos 5.17 e 5.18

Nos gráficos não se observa variações significativas na distribuição do fluxo de pagamentos quando usa-se os parâmetros $V_{Max}=100$ ou $V_{Max}=1000$. Em ambos, as médias do inverso dos índices IHH_{pag} e IHH_{rec} foram, respectivamente, 28,8 e 28,9. As performances dos modelos também não apresentaram diferenças relevantes, conforme mostrado na tabela a seguir.

	$V_{Max}=1000$		$V_{Max}=100$	
	Média	Desvio-Padrão	Média	Desvio-Padrão
LBTR	37,0%	12,7%	36,7%	12,9%
Bech-Soramäki	33,4%	12,3%	33,6%	12,8%
Bilateral Direcionado	88,3%	3,3%	88,5%	4,2%

Tabela 5.4

A comparação com o modelo Bilateral Aleatório é apresentada abaixo.



Gráficos 5.19 e 5.20

O aumento da variância entre os pagamentos não alterou a performance do modelo Bilateral Aleatório que também manteve a performance média de 72,6% em ambos os cenários. A relação entre os tempos de processamento, contudo, aumentou de 0,92 para 1,16 (o desvio-padrão se manteve em 0,16), revelando maior ineficiência relativa do modelo Bilateral em relação ao Bilateral Assistido.

O principal impacto do aumento da dispersão dos pagamentos, contudo, é observado quando se compara os tempos de processamento do **modelo Bilateral Assistido** em ambos os cenários, conforme apresentado no gráfico abaixo.

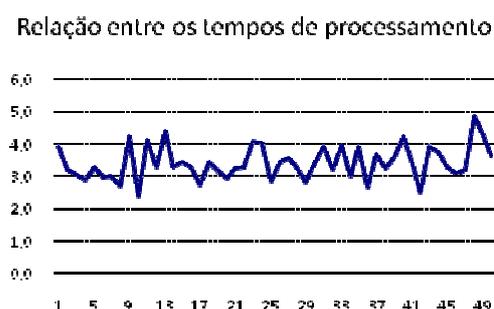


Gráfico 5.21

O tempo de processamento aumentou em média 3,4 vezes (desvio-padrão igual a 0,53), revelando que o aumento da variância entre os pagamentos gera maior esforço computacional para a obtenção de resultados semelhantes. A justificativa desse fato decorre da redução da compensação entre os fluxos de pagamentos com o aumento da dispersão dos mesmos. Esse efeito é observado quando toma-se dois bancos {A,B} que mantenham K pagamentos enfileirados entre si, ou seja: $x_{A,B}=\{x_{A,B,1}, \dots, x_{A,B,K}\}$; $x_{B,A}=\{x_{B,A,1}, \dots, x_{B,A,K}\}$.

Os pagamentos têm distribuição uniforme entre $[L_i, L_s]$ e a variância entre eles será:

$$Var(x_{A,B,k}) = \frac{(L_s - L_i)^2}{12}$$

Por serem gerados de forma independente, a variância da soma dos pagamentos será $K \times Var(x_{A,B,k})$. O efeito do aumento da amplitude dos pagamentos (VMax), conseqüentemente, terá uma relação quadrática na variância da respectiva soma. Para o algoritmo, o impacto do aumento da variância da soma dos pagamentos é direto, ocorrendo através da redução da compensação entre o fluxo bilateral dos bancos. Assim, a maior variância entre a soma dos pagamentos reduz o efeito das respectivas

compensações e aumenta o número de vezes que a liquidez deve percorrer os bancos para liquidar os pagamentos bilaterais.

6. Conclusão:

A liquidação em Sistemas de Pagamentos pode ocorrer em tempo contínuo pelo valor bruto ou ser diferida para apuração do valor líquido. No primeiro caso, o banco central está menos exposto ao risco de inadimplência dos bancos e no segundo há maior eficiência no uso dos recursos. Esta dissertação analisa um modelo híbrido onde os pagamentos são acumulados em filas permitindo a liquidação pelo valor líquido da posição multilateral do participante. Quando os recursos no sistema são insuficientes para liquidar todos os pagamentos existentes, a solução ótima depende da exclusão de alguns pagamentos no processo de liquidação. Esse problema, conhecido na literatura como *Banking Clearing Problem (BCP)*, é NP completo e, portanto, a complexidade para encontrar a solução ótima aumenta exponencialmente com o acréscimo do número de participantes.

O modelo desenvolvido apresenta um critério de mapeamento da rede de pagamentos em função dos preços-sombra do *BCP* por programação linear. Esses valores – λ e μ – mostram, respectivamente, o efeito no valor dos pagamentos liquidados para um incremento marginal nas restrições de liquidez (ou seja, no saldo de recursos no momento da liquidação) e de consistência (que limita a liquidação à fração no intervalo $[0,1]$), e permitem simplificar a busca da solução do problema.

Organizando os bancos em agrupamentos conforme o valor de λ , é possível priorizar a liquidação dos pagamentos que marginalmente possuem maior capacidade de potencializar novas liquidações. O modelo evidencia também que todas as transferências com mesma procedência e destinação têm igual prioridade, permitindo a agregação dos pagamentos para cômputo dos preços-sombra e reduzindo o esforço computacional.

A atribuição de prioridade reduz o esforço para a obtenção do resultado, limitando as combinações existentes para se maximizar a liquidação dos pagamentos existentes.

A seção 4 propôs um algoritmo que avalia a aplicação dos conceitos do modelo. A busca é feita através de sucessivas otimizações bilaterais, priorizando as liquidações de pagamentos que transfiram recursos aos bancos com maior valor λ . Nesse algoritmo, contudo, são tomados os cuidados para evitar empocamento de liquidez e, portanto, dar maior eficiência ao uso da liquidez no processo de liquidação dos pagamentos.

Os resultados do algoritmo foram comparados com outros modelos de liquidação, incluindo um modelo de liquidação bilateral de pagamentos onde os pares dos bancos

são selecionados de forma aleatória. Todos os modelos foram comparados de forma a verificar o desempenho em diferentes cenários, simulando variações: na distribuição dos pagamentos entre os bancos; na distribuição da liquidez; e na amplitude entre o valor máximo e mínimo dos pagamentos.

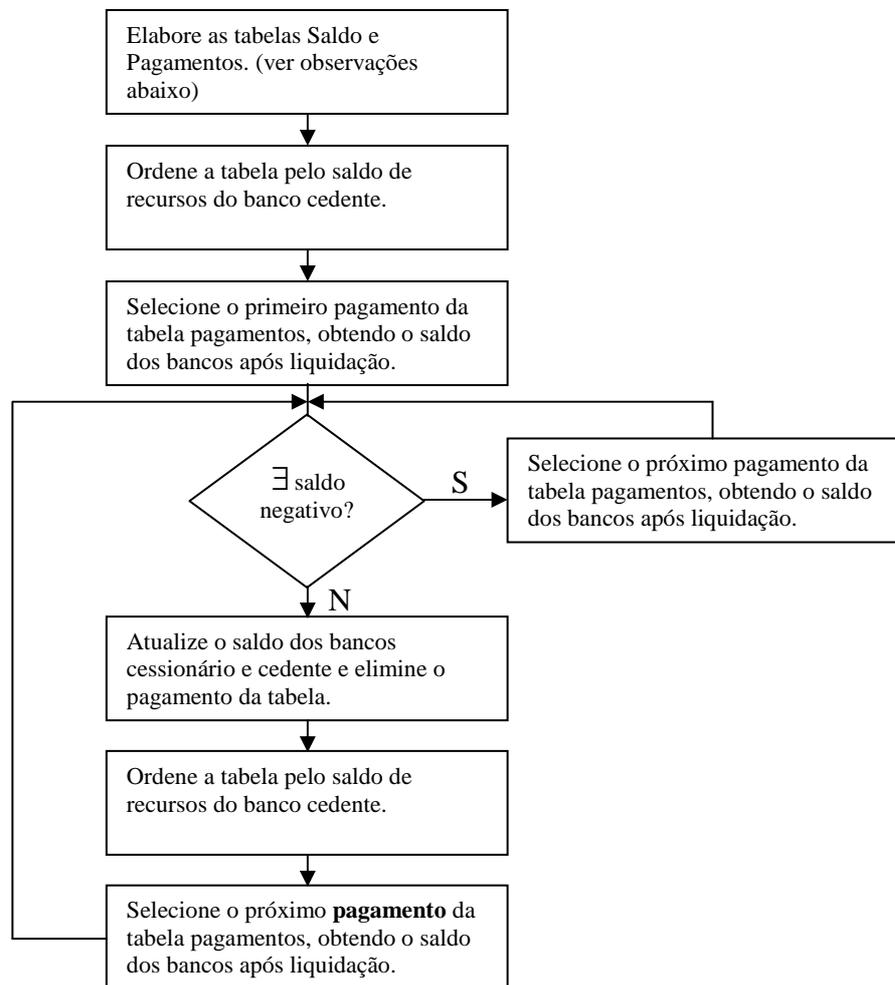
Em todas as simulações verificou-se maior eficiência do algoritmo proposto na dissertação e, na comparação com o modelo bilateral aleatório, foi possível observar diretamente o impacto do uso do critério de priorização na seleção dos pagamentos. Quando a liquidação é ordenada priorizando os pagamentos com maior capacidade marginal de gerar novos pagamentos, observa-se maior volume de transferências efetivadas. Em cenários onde ambos os modelos apresentam valores liquidados equivalentes, o algoritmo da seção 4 apresenta menor tempo de processamento.

Estes resultados evidenciam, de forma robusta, a vantagem da aplicação dos conceitos desenvolvidos sobre os demais modelos de liquidação. A aplicação desses conceitos, contudo, não se limita a otimizações bilaterais – conforme o algoritmo apresentado –, havendo oportunidade para desenvolvimento de algoritmos mais eficientes e criativos na busca da solução ótima. Dessa forma, o modelo elaborado objetiva desvendar algumas características da rede de pagamentos interbancários incentivando novos estudos que não somente aprimorem os conceitos elaborados, mas também sua aplicação.

Bibliografia:

- Angelini, P (1998), “An analysis of competitive externalities in gross settlement systems”, *Journal of Banking & Finance* 22, 1-18.
- Bank for International Settlements, “Real Time Gross Settlement Systems”, Basel, 1997.
- Bandt, O. and Hartmann, P. (2000), “Systemic Risk: A Survey”, Working Paper 35, European Central Bank.
- Bech, M. and Garratt, R. (2003), “The intraday liquidity management game”, *Journal of Economic Theory* 109, 198-219.
- Bech, M. and Soramäki, K. (2001), “Gridlock Resolution in Interbank Payment Systems”, Bank of Finland – Discussion Papers 9/2001 – 34 pages.
- Diamond, D., and Dybvig, P. (1983), “Bank runs, deposit insurance, and liquidity”, *Journal of Political Economy* 91, 401-419.
- Freixas, X. and Parigi, B. (1998), “Contagion and Efficiency in Gross and Net Interbank Payment Systems”, *Journal of Financial Intermediation* 4, 305-327.
- Leinonen, H. (1998), “Interbank funds transfer system: liquidity needs, counterparty risks and collateral”, Bank of Finland – Discussion Papers 16/1998 – 41 pages.
- Lester, B, Millard, S P, and Willison, M (2006), “Optimal settlement rules for payment systems”, Bank of England, mimeo.
- Penalzoa, R (2002), “On Shadow-Price of Banks in Real-Time Gross Settlement Systems”, Universidade de Brasília – Departamento de Economia – Publicações – 20 de dezembro de 2002 – 31 páginas.

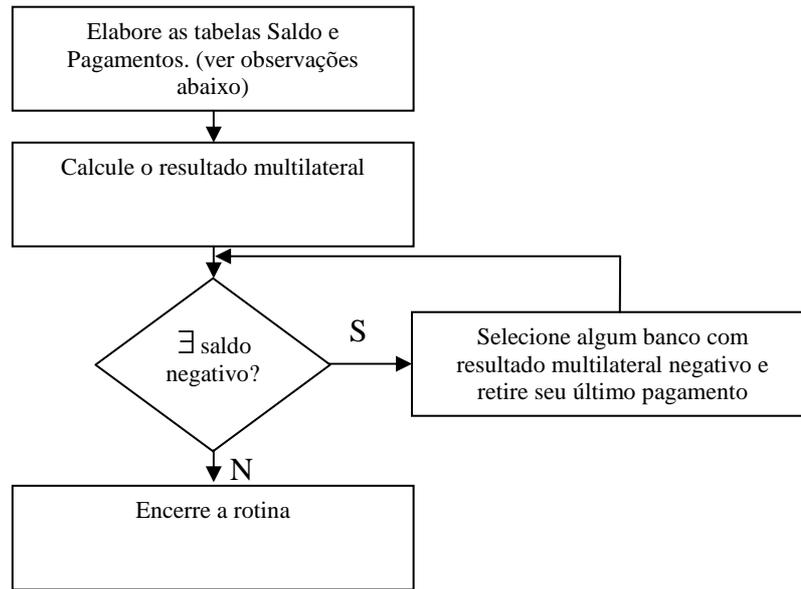
Apêndice 1: Modelo LBTR sem restrição sequencial.



Observação 1: A tabela Saldo exibe o banco e o saldo em recursos para pagamentos em sua conta no banco central.

Observação 2: A tabela Pagamentos desse algoritmo tem quatro colunas: banco cedente, banco cessionário, valor do pagamento e saldo do banco cedente. A quarta coluna é inserida para ordenamento da tabela.

Apêndice 2: Modelo Bech-Soramäki



Observação 1: A tabela Saldo exhibe o banco e o saldo em recursos para pagamentos em sua conta no banco central.

Observação 2: A tabela Pagamentos desse algoritmo tem três colunas: banco cedente, banco cessionário e valor do pagamento.