

Estudo comparativo de algoritmos para computar o portfólio ótimo via Paridade de Risco

Daniel Roizman

Junho 2023

Resumo

Este trabalho procura estimar portfólios ótimos usando a estratégia *Equal Risk Contribution* (ERC), através de métodos numéricos. Para isso, foram selecionados 490 das atuais 500 empresas listadas no índice S&P 500, com intenção de estimar carteiras ótimas com 20, 100 e 490 empresas. Visando evitar o mal condicionamento das matrizes de covariância estimadas, é usada uma seleção dos componentes principais (inspirado no método *Principal Component Analysis* PCA). A ideia é deixar p ativos do conjunto inicial que gerem uma matriz de covariância mais bem condicionada. Feito isso, os métodos numéricos usados para encontrar o portfólio ERC foram os Algoritmos de Newton-Raphson e *Cyclical Coordinate Descent* (CCD). A implementação foi feita em *python*, e foram comparados os tempos de computação dos algoritmos, bem como o número de iterações rodado por estes.

1 Introdução

A escolha ótima dentre um conjunto de ativos tem sido um dos grandes enigmas do mundo das finanças por muito tempo. O cerne do problema reside em tomar decisões sobre o futuro com base unicamente em informações passadas. Nos anos 50 surgiu a Teoria Moderna de Portfolio, desenvolvida por Harry Markowitz (1952), cuja finalidade era resolver esse enigma da alocação eficiente. Inicialmente, uma proposta consistia em otimizar a relação risco-retorno. Para isso, inicialmente, deve haver uma medida de risco. Foi usado amplamente, para tal, o desvio-padrão (volatilidade), de forma que ativos com baixa volatilidade (histórica) e elevados retornos esperados são desejáveis, em detrimento de ativos, por exemplo, com elevados retornos esperados mas com alta volatilidade. A isso se denominou análise de Média-Variância, perspectiva sob a qual busca-se otimizar o portfólio combinando retorno esperado e volatilidade de maneira eficiente. Em outras palavras, um portfólio deve ter seu capital diversificado em alguns ativos que otimizem uma determinada relação risco-retorno desse portfólio. O portfólio *Maximum Sharpe Ratio* (MSR), por exemplo, é definido nesse contexto.

O problema principal dos modelos originais de otimização de Média-Variância talvez seja sua (sensível) dependência no vetor de previsão de retornos esperados. Uma alteração nesse vetor gera grandes alterações no resultado. Uma alternativa ao portfólio MSR, por exemplo, veio a ser o portfólio de Mínima Variância (MV), cujo nome é autoexplicativo. Ambos os modelos padecem do mesmo problema, que é a concentração do risco do portfólio em poucos ativos.

Uma das mais populares opções ao portfólio MSR e MV foi a Paridade de Risco. A primeira grande vantagem dessa nova perspectiva é não precisar estimar retornos esperados, mas apenas o risco dos ativos (e consequentemente, sua covariância). A segunda, pode-se dizer, é a maior estabilidade (numérica) das soluções. Os modelos de Paridade de Risco buscam, em vez de alocar frações do capital, alocar “frações do risco”, diversificando os riscos do portfólio. Ou seja, dado um nível desejável de diversificação, se quer que a contribuição de risco dos ativos ao risco do portfólio obedeça a um vetor de livre escolha. Pode-se argumentar que o portfólio ERC é semelhante ao de Mínima Variância, porém com a condição de diversificação.

Na ótica da Paridade de Risco, o portfólio *Equal Risk Contribution* (ERC) busca dividir igualmente, dentre os ativos, o risco total. A dificuldade desse método consiste em que, para calcular o risco do portfólio, é preciso o peso (percentual) de cada ativo, mas para calcular o peso ótimo de cada ativo, deve-se calcular o risco do portfólio. Ou seja, salvo casos específicos, o problema é endógeno pois a solução acaba sendo autorreferencial. Isso justifica a necessidade de métodos numéricos.

O problema dos métodos numéricos, contudo, pode ser a instabilidade dos resultados. Neste trabalho, para reduzir esse risco, as matrizes de covariância passaram por um processo de “limpeza” dos dados. Esse processo é semelhante a Análise de Componentes Principais, mas simplificada. Consiste em calcular o posto efetivo usando a decomposição SVD, para filtrar os p maiores valores singulares. A razão justificativa é que matrizes grandes, comumente, têm mais chance de ter autovalores discrepantes, o que deixa a matriz mais sensível a erros de estimação, daí a necessidade de melhor condicioná-la para dar mais estabilidade aos resultados. Ou seja, foram filtradas as porções mais bem condicionadas da matriz original, para ter melhores possibilidades de composição do portfólio ERC.

Por fim, vale ressaltar que o objetivo deste trabalho é estudar e comparar os algoritmos, bem como suas performances, e não testar a eficiência financeira dos métodos de Paridade de Risco. Ou seja, não são realizados backtests, comparações com outras estratégias nem nada do tipo.

2 Problema do portfólio ERC

2.1 Definições, Medidas de Risco e Coerência

Em primeiro lugar, para resolver qualquer problema de minimização de risco de um ativo, deve-se definir “risco” (a princípio poderia ser a variância, *value-at-risk*, *expected shortfall* etc.). Para isso, Artzner (1999) define como “medida coerente de risco” as medidas de risco $R: \mathbb{R}^n \rightarrow \mathbb{R}$, em função de um ativo (ou portfólio de ativos) x que seguem às propriedades a seguir:

- i. Subaditividade: $R(x_1 + x_2) = R(x_1) + R(x_2)$;
- ii. Homogeneidade: $R(ax) = aR(x)$;
- iii. Monotonicidade: $x_1 < x_2 \Rightarrow R(x_1) < R(x_2)$;

- iv. Invariância Translacional: $R(x + m) = R(x) - m, \forall m \in \mathbb{R}$ (adicionar “caixa” ao portfólio reduz o risco);

Embora imperfeita, essa definição recebeu tremenda importância do mundo das finanças quantitativas nas últimas décadas. Nesse sentido, sendo a volatilidade (desvio-padrão), uma medida coerente de risco, será a medida utilizada neste trabalho, e não cabe maior profundidade sobre o tema. A partir de agora, quando for dito “risco de um ativo (ou portfólio)”, será a volatilidade. Além disso, outra propriedade fundamental do risco é sua convexidade. O risco de um portfólio com dois ativos é igual ou menor a somas dos riscos correspondentes. A convexidade é necessária à identidade de Euler, a seguir.

Vale definir, agora, um novo conceito que deriva do anterior. A cada ativo corresponde um risco individual σ_i , e um peso w_i no portfólio, que determinarão o quanto cada ativo contribuirá ao risco total do portfólio, $R(\mathbf{w})$. Sendo \mathbf{C} a matriz simétrica de covariância, a volatilidade do portfólio se dá por:

$$R(\mathbf{w}) = \sqrt{\mathbf{w}^T \mathbf{C} \mathbf{w}}$$

De onde se obtém que as contribuições individuais de risco RC_i são:

$$RC_i = w_i \frac{\partial R(\mathbf{w})}{\partial w_i} = \frac{w_i (\mathbf{C} \mathbf{w})_i}{\sqrt{\mathbf{w}^T \mathbf{C} \mathbf{w}}}$$

Dada a homogeneidade e convexidade da função $R(\mathbf{w})$, a Identidade de Euler nos garante que:

$$R(\mathbf{w}) = \sum_{i=1}^n w_i \frac{\partial R(\mathbf{w})}{\partial w_i} = \sum_{i=1}^n RC_i = \sum_{i=1}^n \frac{w_i (\mathbf{C} \mathbf{w})_i}{\sqrt{\mathbf{w}^T \mathbf{C} \mathbf{w}}}$$

Note que a parte direita (das três igualdades acima) só ocorre quando a medida de risco escolhida é a volatilidade, podendo não ser válida para outras medidas.

Nesse contexto, os diversos modelos de otimização de portfólio buscam, muitas vezes, determinar o vetor $\mathbf{w} \in \mathbb{R}^n$ de pesos de ativos, sujeito a restrições (que variam conforme a intenção do investidor). Neste trabalho, as restrições serão que a soma dos pesos dos ativos totalize 100%; que todos os pesos sejam positivos (*long only*); e que o vetor de contribuições totais de risco dos ativos convirja ao vetor pré-especificado \mathbf{b} , cujos elementos também devem somar 100%. Mais especificamente, será escolhido o vetor \mathbf{b} de pesos iguais ($b_i = \frac{1}{n}$, $i \in$

$1, \dots, n$), não com a intenção de alocar percentuais do capital dentre os ativos, mas de alocar percentuais do risco, e a partir daí determinar a alocação de capital.

Noutras palavras, o portfólio ERC requer vetores \mathbf{w} de pesos, compostos por n ativos, cujas contribuições (iguais) ao risco do portfólio sejam $RC_i = RC_j$, $\forall i, j$. O problema, então, é descrito pelo seguinte sistema:

$$\left\{ \begin{array}{l} RC_i = b_i R(\mathbf{w}) \\ b_i > 0 \\ w_i \geq 0 \\ \sum_{i=1}^n b_i = 1 \\ \sum_{i=1}^n w_i = 1 \end{array} \right.$$

No caso do portfólio ERC, como já dito, pode-se usar $1/n$ no lugar de b_i . A princípio, têm-se um sistema não-linear com restrições. Como esse sistema carece de solução analítica, opta-se pelo uso de métodos numéricos, de modo que o sistema pode ser reescrito como um problema de otimização da seguinte maneira:

$$\mathbf{w}^* = \arg \min f(\mathbf{w}, \mathbf{b})$$

$$\text{Restrito a } \mathbf{1}^T \mathbf{w} = 1, \quad 0 \leq \mathbf{w} \leq 1$$

No entanto, a formulação da função objetivo f pode variar conforme o método usado na resolução do problema. A função objetivo depende do método a ser usado. Neste trabalho, assim como Roncalli (2019), se incluirá na formulação do problema de otimização uma barreira logarítmica:

$$\mathbf{w}^* = \arg \min R(\mathbf{w})$$

$$\text{Restrito a } \left\{ \begin{array}{l} \sum_{i=1}^n b_i \ln(w_i) \geq K^* \\ \mathbf{1}^T \mathbf{w} = 1 \\ 0 \leq \mathbf{w} \leq 1 \end{array} \right.$$

onde K^* é uma constante a ser determinada, que age como um coeficiente de diversificação do portfólio, pois um vetor de pesos \mathbf{w} mais concentrado possui menor $\sum_{i=1}^n b_i \ln(w_i)$. A ideia por trás dos métodos numéricos será tentar encontrar aproximações lineares a esse problema. Antes de detalhar os métodos

utilizados, é válido comentar algumas propriedades do portfólio ERC. Daí se considera o seguinte Lagrangiano:

$$\mathcal{L}(\mathbf{w}) = \frac{1}{2} \mathbf{w}^T \mathbf{C} \mathbf{w} - \lambda_c \sum \ln(w_i)$$

e, dependendo de cada método, a otimização da função objetivo acima ocorre de maneiras distintas.

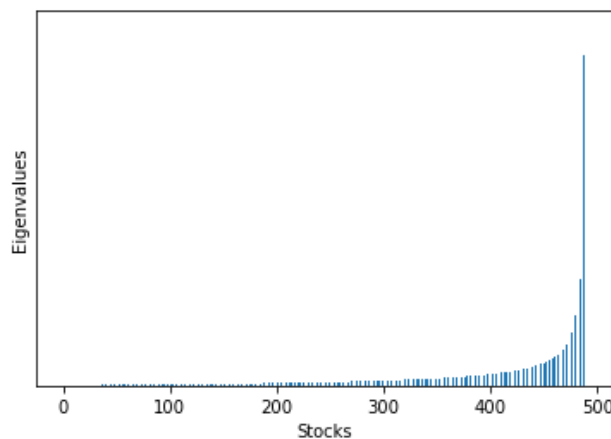
2.2 Existência e Unicidade de Solução

Roncalli (2009) demonstrou que há solução analítica para o portfólio ERC se as correlações forem iguais e constantes. Contudo, mesmo não sendo o caso, a solução existe e é única. A unicidade depende de que o vetor \mathbf{b} seja estritamente positivo. Nesse caso, a solução será um vetor \mathbf{v}^* , que não necessariamente possui componentes que somem 100%, de modo que pode ser feita a seguinte normalização $\mathbf{w}^* = \mathbf{v}^* / \sum v_i$.

2.3 Posto Efetivo da matriz de covariância

A figura abaixo ilustra a discrepância dos autovalores associados a cada variável da matriz de covariância inicial, de tamanho 490×490 . Por isso, é interessante saber quais ativos podem ser excluídos do conjunto inicial. Alguns artigos sugerem métodos de *shrinkage* na estimação da matriz de covariância por diferentes métodos, como Análise de Componentes Principais, e até mesmo uma abordagem baseada no Teorema dos Autovalores Entrelaçados, de Cauchy, como é feito por Fernandes et al (2020).

Dessa forma, antes de escolher o portfólio ótimo, foi feita uma filtragem dentre esses $n = 490$ ativos, ou seja, foram mantidos r ativos dos 490, sendo $p = n - r$ o número de ativos eliminados, e r o posto efetivo da matriz.



1 Autovalores da matriz de covariância original

O cálculo do posto efetivo foi feito usando decomposição SVD, e escolhendo as variáveis (ativos) com os valores singulares maiores que um certo erro [ver Tabela 1].

3 Métodos Numéricos a Serem testados

Antes de apresentar os métodos, é válido explicar seus critérios de convergência. Os processos de iteração dos algoritmos são repetidos até que a convergência. No caso do CCD,

3.1 Newton-Raphson (NR)

De maneira similar ao Método de Newton original (para polinômios), o algoritmo NR multidimensional se trata de linearizar um problema não linear, usando sua aproximação pela Expansão de Taylor, e resolver de maneira iterativa. Dos principais problemas do algoritmo, pode-se destacar que a computação do inverso do Jacobiano é o principal, já que é a etapa mais demorada do algoritmo NR, e pode sofrer com o mal condicionamento da matriz de correlação (por isso a importância de selecionar a submatriz com menor número de condição), principalmente para matrizes grandes.

No problema $\mathbf{w}^* = \arg \min R(\mathbf{w})$, o algoritmo NR consiste em iterar a equação $\mathbf{w}_{k+1} = \mathbf{w}_k + \lambda_k \mathbf{d}_k$ [ver seção 7.2 do Apêndice]. Aqui se define a convergência como $\max |\nabla G(\mathbf{w}^{(k)})_i| < tol$, ou seja, o componente do gradiente com maior valor absoluto deve ser menor que uma certa tolerância. O algoritmo, portanto, é descrito da seguinte forma:

Algoritmo 1 Método de Newton-Raphson

1. $k = 0$
 2. $\max_iter = M$
 3. $tol = t$
 4. **input** $C, \mathbf{w}^{(0)}$
 5. **repeat**:
 6. $\mathbf{w}_{k+1} = \mathbf{w}_k + \lambda_k \mathbf{d}_k$
 7. $k = k + 1$
 8. **until** $\max |\nabla G(\mathbf{w}^{(k)})_i| < tol$ **or** $k > \max_iter$
 9. $\mathbf{w} = \mathbf{w}^{(k)}$
 10. **return** \mathbf{w}
-

O código para o algoritmo NR foi escrito em python, pondo $M = 20$, e $\mathbf{w}^{(0)} = \boldsymbol{\sigma}^{-1} / \|\boldsymbol{\sigma}^{-1}\|_1$, $\boldsymbol{\sigma}^{-1} = \sigma_1^{-1}, \sigma_2^{-1}, \dots, \sigma_n^{-1}$, sendo σ_i a volatilidade do ativo i . A razão da escolha se dá a que esta seria a solução analítica no caso de correlações idênticas, sendo esta a melhor escolha inicial segundo Roncalli (2014).

3.2 Cyclical Coordinate Descent (CCD)

O algoritmo CCD é um método iterativo que busca o vetor \mathbf{w}^* na direção que otimiza a função objetivo. Nesse sentido, se assemelha um pouco ao método dos Gradientes Conjugados, com a (grande) diferença de que, nas iterações do CCD, calcula-se componente a componente do vetor \mathbf{w} , como no método de Gauss-Seidel, por exemplo. Ou seja, se usa no cálculo das componentes $w_i^{(k)}$ – onde k é o índice da iteração – todas as componentes recém calculadas $w_j^{(k)}, j = 1, \dots, i - 1$, junto às componentes “antigas” $w_l^{(k-1)}, l = i + 1, \dots, n$. a convergência ocorre quando $\max |w_i(\mathbf{C}\mathbf{w})_i - b_i| < tol$, que é o mesmo que estabelecer uma tolerância máxima para maior a diferença entre o peso calculado e o peso desejado do i -ésimo ativo. O algoritmo é:

Algoritmo 3 Cyclical Coordinate Descent (CCD)

```

1.  $k = 0$ 
2.  $\text{max\_iter} = M$ 
3. input  $\mathbf{C}$ 
4. initialize  $\mathbf{w}^{(0)}$ 
5. repeat:
6.   for  $i$  in  $1:n$  :
7.      $w_i^{(k+1)} = \arg \min_v f(w_1^{(k+1)}, \dots, w_{i-1}^{(k+1)}, v, w_{i+1}^{(k)}, \dots, w_n^{(k)})$ 
8.   end for
9.    $k = k + 1$ 
10. until convergence or  $k > \text{max\_iter}$ 
11.  $\mathbf{w} = \mathbf{w}^{(k)}$ 
12. return  $\mathbf{w}$ 

```

4 Resultados

Para o algoritmo NR foi escrito o código completo em python, enquanto o código para o CCD estava já disponível na biblioteca *pyfeng*, criada por Choi (2022). Os tempos médios de computação, bem como o número de iterações necessário, para as matrizes de covariância utilizadas é visto na tabela abaixo:

Número de ativos (posto efetivo)	Mínimo dos valores singulares	Newton-Raphson		<i>Cyclical Coordinate Descent</i>	
		Tempo médio (s)	Iterações	Tempo médio (s)	Iterações
490	0,0000	0,0533	5	0,0156	4
360	e^{-10}	0,0156	5	0,0221	4
224	e^{-9}	0,0049	5	0,0156	3
98	e^{-8}	0,0009	5	0,0000	3

Tabela 1 Resultados Computacionais

5 Conclusão e Trabalhos Futuros

Foram apresentados dois dos principais algoritmos usados no cálculo do portfólio ótimo sob a perspectiva da Paridade de Risco. É fácil ver que o algoritmo CCD foi mais eficiente em termos de iterações, mas pode ser mais lento, dependendo do número de ativos. Alguns autores defendem que o CCD seja mais rápido que NR apenas para um número grande de ativos. Nos dois algoritmos apresentados há vantagens e desvantagens, numericamente falando. Para trabalhos futuros, há algumas possibilidades, como aplicar outros métodos de *shrinkage* na estimação da matriz de covariância, ou usar algoritmos mais eficientes pois, além daqueles vistos, há outros, inclusive melhores versões do NR e CCD – como proposto por Choi (2022).

6 Referências

- [1] Artzner, P. Delbarn, F., Eber, J. M., Heath, D. (1999). *Coherent measures of risk*. **Mathematical Finance**. DOI: <https://doi.org/10.1111%2F1467-9965.00068>
- [2] Choi, J., & Chen, R. (2022). *Improved iterative methods for solving risk parity portfolio*. **Journal of Derivatives and Quantitative Studies**, 30(2). DOI: <https://doi.org/10.1108/JDQS-12-2021-0031>
- [3] Fernandes, F., Oliveira, R., De-Losso, R., Jonathan Diaz Soto, A., Delano Cavalcanti, P., Moreira da Silva Campos, G. (2020). *Saving Markowitz: A Risk Parity Approach Based on the Cauchy Interlacing Theorem*. DOI: <http://dx.doi.org/10.2139/ssrn.3654300>
- [4] Markowitz H. (1952), *Portfolio Selection*, **Journal of Finance**, JSTOR: <http://www.jstor.org/stable/2975974?origin=JSTOR-pdf>

- [5] Richard, J. C., Roncalli, T. (2019). *Constrained Risk Budgeting Portfolios: Theory, Algorithms, Applications & Puzzles*. SSRN: <https://ssrn.com/abstract=3331184>
- [6] Roncalli, T. (2014). *Introduction to Risk Parity and Budgeting*. **Chapman & Hall/CRC FINANCIAL MATHEMATICS SERIES**.
- [7] Roncalli, T. (2009). *On the Properties of equally-weighted risk contributions portfolios*. **The Journal of Portfolio Management**. DOI: <https://doi.org/10.3905/jpm.2010.36.4.060>

7 Apêndice

7.1 Derivação do Método de Newton-Raphson

Voltando ao problema $\mathbf{w}^* = \arg \min f(\mathbf{w})$, veja que a condição de otimização deve ocorrer quando $G(\mathbf{w}^*) = 0$, em que $G(\mathbf{w})$ é o gradiente de f , e sua aproximação pela expansão de Taylor de primeira ordem é:

$$G(\mathbf{w}) = G(\mathbf{w}_0) + H(\mathbf{w}_0)(\mathbf{w} - \mathbf{w}_0)$$

sendo H a Hessiana de $f(\mathbf{w})$, de modo que o vetor ótimo \mathbf{w}^* resulta em:

$$\begin{aligned} G(\mathbf{w}^*) = 0 &\Rightarrow G(\mathbf{w}_0) + H(\mathbf{w}_0)(\mathbf{w} - \mathbf{w}_0) = 0 \Rightarrow H(\mathbf{w}_0)\mathbf{w} = H(\mathbf{w}_0)\mathbf{w}_0 - G(\mathbf{w}_0) \\ &\Rightarrow \mathbf{w} = \mathbf{w}_0 - H^{-1}(\mathbf{w}_0)G(\mathbf{w}_0) \end{aligned}$$

Na forma recursiva, passa a ser:

$$\mathbf{w}_{k+1} = \mathbf{w}_k - H^{-1}(\mathbf{w}_k)G(\mathbf{w}_k)$$

Muitos autores preferem, escrever $\mathbf{w}_{k+1} = \mathbf{w}_k + \lambda_k d_k$, onde $d_k = -H^{-1}(\mathbf{w}_k)G(\mathbf{w}_k)$, e é incluído um “passo” $\lambda_k > 0$, já que $\lambda_k = 1$ nem sempre é ótimo (pode não reduzir f). Considerando que a função de Lagrange é:

$$\mathcal{L}(\mathbf{w}) = \frac{1}{2} \mathbf{w}^T \mathbf{C} \mathbf{w} - \lambda_c \sum_{i=1}^n \ln(w_i)$$

obtem-se:

$$G(\mathbf{w}) = \mathbf{C} \mathbf{w} - \lambda_c \mathbf{w}^{-1}, \quad \mathbf{w}^{-1} = (w_1^{-1}, w_2^{-1}, \dots, w_n^{-1})$$

$$H(\mathbf{w}) = \mathbf{C} - \lambda_c \text{diag}(\mathbf{w}^{-2}), \quad \text{diag}(\mathbf{w}^{-2}) = \begin{bmatrix} w_1^{-2} & 0 & \dots & 0 \\ 0 & w_2^{-2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & w_n^{-2} \end{bmatrix}$$

No ponto de ótimo, lembre-se que $\lambda_c = R(\mathbf{w}^*)/n$, então uma boa escolha inicial é $\lambda_c = R(\mathbf{w}^{(0)})/n$, segundo Roncalli (2014).

7.2 Derivação do método CCP

Usando o mesmo Lagrangiano $\mathcal{L}(\mathbf{w})$ do método anterior, a condição de primeira ordem fornece:

$$\frac{\partial \mathcal{L}(\mathbf{w}, \lambda_c)}{\partial \mathbf{w}} = (\mathbf{C}\mathbf{w})_i - \frac{\lambda_c}{w_i} = 0 \Rightarrow w_i(\mathbf{C}\mathbf{w})_i - \lambda_c = 0$$

Daí, segue-se que, sendo $(\mathbf{C}\mathbf{w})_i = \sum_{j=1}^n c_{i,j}w_j$, então, escrevendo $c_{i,i} = \sigma_i^2$, e separando-os do somatório, pode-se escrever:

$$\sigma_i^2 w_i^2 + \left(\sum_{j \neq i} c_{ij}w_j \right) w_i - \lambda_c = 0$$

Observe que se forma uma equação de segundo grau, do tipo $\alpha_i w^2 + \beta_i w + \gamma_i = 0$. A função é convexa pois $\alpha_i = \sigma_i^2 > 0, \forall i = 1, \dots, n$, então o ponto crítico é um ponto de mínimo. Além disso, pondo $\alpha_i = \sigma_i^2, \beta_i = (\sum_{k \neq i} c_{i,k}w_k), \gamma_i = -\lambda_c$, o discriminante $\Delta = \beta_i^2 - 4\alpha_i\gamma_i$ é positivo, então pode-se afirmar que há duas soluções, das quais será escolhida a positiva. Então, sendo (k) o índice da iteração, e pondo:

$$\begin{cases} \alpha_i^{(k+1)} = \sigma_i^2 \\ \beta_i^{(k+1)} = \left(\sum_{j < i} c_{ij}w_j^{(k+1)} + \sum_{j > i} c_{ij}w_j^{(k)} \right) \\ \gamma_i^{(k+1)} = -\lambda_c \end{cases}$$

têm-se um processo iterativo da seguinte forma:

$$w_i^{(k+1)} = \frac{-\beta_i^{(k+1)} \pm \sqrt{\left(\beta_i^{(k+1)}\right)^2 - 4\alpha_i^{(k+1)}\gamma_i^{(k+1)}}}{2\alpha_i^{(k+1)}}$$

Dessa maneira, o algoritmo CCD consiste em iterar a equação acima até a convergência, definida como $\max |w_i(\mathbf{C}\mathbf{w})_i - b_i| < tol$.