KULLBACK-LEIBLER SIMPLEX

Abstract. This technical reference presents the functional structure and the algorithmic implementation of KL (Kullback-Leibler) simplex. It details the simplex approximation and fusion. The KL simplex is fundamental, robust, adaptive an informatics agent for computational research in economics, finance, game and mechanism. From this perspective the study provides comprehensive results to facilitate future work in such areas.

God does not care about our mathematical difficulties. He integrates empirically. Albert Einstein

There is nothing free, except the grace of God. True Grit (2010)

1. Introduction

This paper presents an alternative for sequential optimizing agent which is crucial for the reliability of computational economics research. In particular it is a version of online classifier, a machine learning which processes classification with data stream. The sequential implementation makes it efficient, fast and practical data flow processing. Among this type of classifier, informatics divergence approach stands out with solid foundation in mathematical statistics and informatics theory. It is instructive to see the difference of the two approaches. Standard approach targets the performance in objective function, while the informatics works with statistical measures, e.g. Kullback-Leibler and Renyi divergence [CDR07]. Positively the informatics agent can be effective alternative to standard sequential optimizers.

Furthermore informatics approach delivers powerful concepts, e.g. (i) the advance will leverage the notion and insight from dynamic programming [Sni10]; when a control is simplex and transition matrix, it has a strong foundation in probability and Markov chain [Beh00]. (ii) model-free or agnostic data makes it capable of deriving superior second-order perceptron working the real-world data [BCG05]. This approach consequently can improve machine learning that is robust and applicable for computational research in economics, finance, game and mechanism.

The next section lists useful formula and identity. Section 3 presents the structure of online machine learning [CDF08, LHZG11] and key results; section 4 discusses the implementation. The instructive remarks are in section 5 and the proof is in Appendix.

2. The matrix

Simplex [CY11].

$\langle 1\rangle\, \boldsymbol{\mu} \in \overset{\leftrightarrow}{\triangle} \Leftrightarrow \boldsymbol{\mu}\cdot\mathbf{1} = \mathbf{1}$ and $\langle 2\rangle\, \boldsymbol{\mu} \in \triangle \Leftrightarrow \boldsymbol{\mu} \in \overset{\leftrightarrow}{\triangle}$, with $\min(\boldsymbol{\mu}) \geq 0$.

Taylor expansion. $\ln(\boldsymbol{\mu}\cdot\boldsymbol{x}_i) \approx \ln(\boldsymbol{\mu}_i\cdot\boldsymbol{x}_i) + \frac{(\boldsymbol{\mu}-\boldsymbol{\mu}_i)\cdot\boldsymbol{x}_i}{\boldsymbol{\mu}_i\cdot\boldsymbol{x}_i}$.

Date: November 22, 2011.

Key words and phrases. KL divergence, second-order perceptron, informatics agent, simplex projection and fusion.

Acknowledgments. Economics department at Thammasat and Queen's university greatly supports the study. The author warmly thanks Frank Flatters, Frank Milne, Ted Neave and Pranee Tinakorn for the encouragement, without which this paper would not had been written. He certainly appreciates comments and discussions with the Sukniyom, Ko-Kao, Chai-Shane, Lek-Air, Ron-Fran, Pui, NaPoj and of course, Kay and Nongyao.

Contact. facebook.com/popon.kangpenkae.

Symmetric squared decomposition (SSD). $\Sigma_{[i]} = \Upsilon_{[i]}^2, \Upsilon_{[i]} = Q_{[i]}\sqrt{\operatorname{diag}\left(\lambda_{[i,]1}, \ldots, \lambda_{[i,]d}\right)}Q_{[i]}^\top : Q_{[i]}$ is orthogonal and the eigenvector of $\Sigma_{[i]}$; $\left(\lambda_{[i,]1}, \ldots, \lambda_{[i,]d}\right)$ is the eigenvalue of $\Sigma_{[i]}$. Of course $\Upsilon_{[i]}, \Sigma_{[i]}$ is symmetric PSD.

Inversion [PP08] [146].

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B\left(C^{-1} + DA^{-1}B\right)^{-1} DA^{-1}$$

for our application,

$$\Sigma^{-1} = \Sigma_i^{-1} + \frac{\boldsymbol{x}_i\boldsymbol{x}_i^\top}{c} \Rightarrow \Sigma = \Sigma_i - \frac{\Sigma_i\boldsymbol{x}_i\boldsymbol{x}_i^\top\Sigma_i}{c + \boldsymbol{x}_i^\top\Sigma_i\boldsymbol{x}_i}$$

Differentiation [PP08] [78, 49, 102, 83].

$\frac{\partial}{\partial\boldsymbol{\mu}}\left(\boldsymbol{\mu}_i - \boldsymbol{\mu}\right)\Upsilon_i^{-2}\left(\boldsymbol{\mu}_i - \boldsymbol{\mu}\right) = 2\Upsilon_i^{-2}\left(\boldsymbol{\mu} - \boldsymbol{\mu}_i\right)$

$\frac{\partial}{\partial\Upsilon}\ln\left(\det\Upsilon^2\right) = 2\Upsilon^{-1}$

$\frac{\partial}{\partial\Upsilon}\operatorname{Tr}\left(\Upsilon_i^{-2}\Upsilon^2\right) = \Upsilon_i^{-2}\Upsilon + \Upsilon\Upsilon_i^{-2}$

$\frac{\partial}{\partial\Upsilon}\boldsymbol{x}_i^\top\Upsilon^2\boldsymbol{x}_i = \boldsymbol{x}_i\boldsymbol{x}_i^\top\Upsilon + \Upsilon\boldsymbol{x}_i\boldsymbol{x}_i^\top = \frac{\partial}{\partial\Upsilon}\left\|\Upsilon\boldsymbol{x}_i\right\|^2$

$\frac{\partial}{\partial\Upsilon}\left\|\Upsilon\boldsymbol{x}_i\right\| = \frac{\boldsymbol{x}_i\boldsymbol{x}_i^\top\Upsilon + \Upsilon\boldsymbol{x}_i\boldsymbol{x}_i^\top}{2\|\Upsilon\boldsymbol{x}_i\|} = \frac{\boldsymbol{x}_i\boldsymbol{x}_i^\top\Upsilon + \Upsilon\boldsymbol{x}_i\boldsymbol{x}_i^\top}{2\sqrt{\boldsymbol{x}_i^\top\Upsilon^2\boldsymbol{x}_i}}$

KL divergence.

$$D_{KL}\left(\mathbb{N}\left(\boldsymbol{\mu}, \Upsilon^2\right) || \mathbb{N}\left(\boldsymbol{\mu}_i, \Upsilon_i^2\right)\right) = \frac{1}{2}\left[\ln\left(\frac{\det\Upsilon_i^2}{\det\Upsilon^2}\right) + \operatorname{Tr}\left(\Upsilon_i^{-2}\Upsilon^2\right) + \left(\boldsymbol{\mu}_i - \boldsymbol{\mu}\right)\Upsilon_i^{-2}\left(\boldsymbol{\mu}_i - \boldsymbol{\mu}\right)\right]$$

3. Approximation

3.1. This section refers to [CDR07, LHZG11] for the model concept and definition.

As KL simplex solution in $\triangle$ does not have a closed form, the approximation will start with $\overset{\leftrightarrow}{\triangle}$,

$$\left(\boldsymbol{\mu}_{i+1}, \Sigma_{i+1}\right) = \arg\min D_{KL}\left(\mathbb{N}\left(\boldsymbol{\mu}, \Sigma\right) || \mathbb{N}\left(\boldsymbol{\mu}_i, \Sigma_i\right)\right)$$

subject to $\hbar\left(y_i f\left(\boldsymbol{\mu}\cdot\boldsymbol{x}_i\right) - \epsilon\right) \geq \phi\sqrt{\boldsymbol{x}_i^\top\Sigma\boldsymbol{x}_i}, y_i \in \{-1, 1\}$, and $\boldsymbol{\mu} \in \overset{\leftrightarrow}{\triangle}$.

Applying the main result in [LCLMV04] $[VI.2]$, an invariance theorem is straightforward,

Theorem. *The optimal pair* $\left(\boldsymbol{\mu}_{i+1}, \Sigma_{i+1}\right)$ *is invariant to similarity-metric divergences.*

We consider $\left[\text{normal, hinge, hinge}^2\right]$ constraint (see section Section 5), with two flavors: $\{\text{linear,logarithm}\} = \{[ln], [\ln]\} \ni f(.)$. Let $\Sigma_{[i]} = \Upsilon_{[i]}^2$ where $\Upsilon_{[i]}$ has SSD, the $\hbar-$Lagrangian is

$$\mathcal{L} = \frac{1}{2}\left[\ln\left(\frac{\det\Upsilon_i^2}{\det\Upsilon^2}\right) + \operatorname{Tr}\left(\Upsilon_i^{-2}\Upsilon^2\right) + \left(\boldsymbol{\mu}_i - \boldsymbol{\mu}\right)\Upsilon_i^{-2}\left(\boldsymbol{\mu}_i - \boldsymbol{\mu}\right)\right] + \alpha\left(\phi\left\|\Upsilon\boldsymbol{x}_i\right\| - \hbar\right) + \rho\left(\boldsymbol{\mu}\cdot\mathbf{1} - 1\right)$$

Define hinge function $\lfloor z\rfloor = \max\{0, z\}$ and $\langle z\rangle = \lfloor z\rfloor / |z| \in \{0, 1\}$.

3.2. [**normal**] , $\hbar_{\emptyset}$.

3.2.1. *Linear* : $\hbar_{\emptyset[ln]}$.

Lemma 1. $\Sigma_{i+1}^{-1} \blacktriangleright \hbar_{\emptyset[ln]}$

$$\Sigma_{i+1}^{-1} = \Sigma_i^{-1} + \alpha\phi \frac{\boldsymbol{x}_i \boldsymbol{x}_i^\top}{\sqrt{\boldsymbol{x}_i^\top \Sigma_{i+1} \boldsymbol{x}_i}}$$

Lemma 2. $\Sigma_{i+1} \blacktriangleright \hbar_{\emptyset[ln]}$

$$\Sigma_{i+1} = \Sigma_i - \beta \Sigma_i \boldsymbol{x}_i \boldsymbol{x}_i^\top \Sigma_i$$

where $\beta = \frac{\alpha\phi}{\sqrt{u_i} + \alpha\phi\upsilon_i}, (u_i, \upsilon_i) \equiv \left(\boldsymbol{x}_i^\top \Sigma_{i+1} \boldsymbol{x}_i, \boldsymbol{x}_i^\top \Sigma_i \boldsymbol{x}_i\right)$.

Lemma 3. $\sqrt{u_i} \blacktriangleright \hbar_{\emptyset[ln]}$

$$\sqrt{u_i} = \frac{-\alpha\phi\upsilon_i + \sqrt{\alpha^2\phi^2\upsilon_i^2 + 4\upsilon_i}}{2}$$

Lemma 4. $\boldsymbol{\mu}_{i+1} \blacktriangleright \hbar_{\emptyset[ln]}$

$$\boldsymbol{\mu}_{i+1} = \boldsymbol{\mu}_i + \alpha y_i \Sigma_i \left(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i\right)$$

where $\overline{\boldsymbol{x}} = \overline{x}\mathbf{1} \equiv \frac{\mathbf{1}^\top \Sigma_i \boldsymbol{x}_i}{\mathbf{1}^\top \Sigma_i \mathbf{1}} \mathbf{1}$

Lemma 5. $\alpha \blacktriangleright \hbar_{\emptyset[ln]}$, $\alpha = \left\lfloor \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \right\rfloor$ *such that*

$(a, b, c) = \left(\lambda' \left(\lambda' + \upsilon_i\phi^2\right), 2\lambda\left(\lambda' + \frac{\upsilon_i\phi^2}{2}\right), \lambda^2 - \upsilon_i\phi^2\right)$
$\left(\lambda, \lambda'\right) = \left(y_i \left(\boldsymbol{\mu}_i \cdot \boldsymbol{x}_i\right) - \epsilon, \boldsymbol{x}_i^\top \Sigma_i \left(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i\right)\right)$

3.2.2. *Logarithm* : $\hbar_{\emptyset[\mathrm{ln}]}$.

Lemma 6. $\Sigma_{i+1}^{-1} \blacktriangleright \hbar_{\emptyset[\mathrm{ln}]} \equiv$ Lemma 1.

Lemma 7. $\Sigma_{i+1} \blacktriangleright \hbar_{\emptyset[\mathrm{ln}]} \equiv$ Lemma 2.

Lemma 8. $\sqrt{u_i} \blacktriangleright \hbar_{\emptyset[\mathrm{ln}]} \equiv$ Lemma 3.

Lemma 9. $\boldsymbol{\mu}_{i+1} \blacktriangleright \hbar_{\emptyset[\mathrm{ln}]}$

$$\boldsymbol{\mu}_{i+1} \approx \boldsymbol{\mu}_i + \frac{\alpha y_i}{\boldsymbol{\mu}_i \cdot \boldsymbol{x}_i} \Sigma_i \left(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i\right),$$

where $\overline{\boldsymbol{x}} = \overline{x}\mathbf{1} \equiv \frac{\mathbf{1}^\top \Sigma_i \boldsymbol{x}_i}{\mathbf{1}^\top \Sigma_i \mathbf{1}} \mathbf{1}$.

Lemma 10. $\alpha \blacktriangleright \hbar_{\emptyset[\mathrm{ln}]}$, $\alpha = \left\lfloor \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \right\rfloor$ *such that*

$(a, b, c) = \left(\lambda' \left(\lambda' + \upsilon_i\phi^2\right), 2\lambda\left(\lambda' + \frac{\upsilon_i\phi^2}{2}\right), \lambda^2 - \upsilon_i\phi^2\right)$
$\left(\lambda, \lambda'\right) \approx \left(y_i \ln\left(\boldsymbol{\mu}_i \cdot \boldsymbol{x}_i\right) - \epsilon, \frac{\boldsymbol{x}_i^\top \Sigma_i \left(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i\right)}{\left(\boldsymbol{\mu}_i \cdot \boldsymbol{x}_i\right)^2}\right)$

3.3. $[\textbf{hinge}], \hbar_1$ and $\left[\textbf{hinge}^2\right], \hbar_2$.

3.3.1. *Linear* : $\hbar_{1[ln]}, \hbar_{2[ln]}$.

$\Sigma_{i+1}^{-1} \blacktriangleright \hbar_{[1,2][ln]}$, Lemma 11 $\equiv$ Lemma 21 $\equiv$ Lemma 1, $\Sigma_{i+1}^{-1} \blacktriangleright \hbar_{\emptyset[ln]}$

$\Sigma_{i+1} \blacktriangleright \hbar_{[1,2][ln]}$, Lemma 12 $\equiv$ Lemma 22 $\equiv$ Lemma 2, $\Sigma_{i+1} \blacktriangleright \hbar_{\emptyset[ln]}$

$\sqrt{u_i} \blacktriangleright \hbar_{[1,2][ln]}$, Lemma 13 $\equiv$ Lemma 23 $\equiv$ Lemma 3, $\sqrt{u_i} \blacktriangleright \hbar_{\emptyset[ln]}$

Lemma 14. $\boldsymbol{\mu}_{i+1} \blacktriangleright \hbar_{1[ln]}$

$$\boldsymbol{\mu}_{i+1} = \boldsymbol{\mu}_i + \langle y_i (\boldsymbol{\mu}_i \cdot \boldsymbol{x}_i) - \epsilon \rangle \alpha y_i \Sigma_i (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i)$$

where $\overline{\boldsymbol{x}}_i = \overline{x}_i \mathbf{1} \equiv \frac{\mathbf{1}^\top \Sigma_i \boldsymbol{x}_i}{\mathbf{1}^\top \Sigma_i \mathbf{1}} \mathbf{1}$

Lemma 15. $\alpha \blacktriangleright \hbar_{1[ln]}$, $\alpha = \left\lfloor \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \right\rfloor$ *such that*

$(a, b, c) = \left(\lambda' \left(\lambda' + \upsilon_i \phi^2\right), 2\lambda \left(\lambda' + \frac{\upsilon_i \phi^2}{2}\right), \lambda^2 - \upsilon_i \phi^2\right)$,
$\left(\lambda, \lambda'\right) = \left(y_i (\boldsymbol{\mu}_i \cdot \boldsymbol{x}_i) - \epsilon, \boldsymbol{x}_i^\top \Sigma_i (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i)\right)$

Lemma 24. $\boldsymbol{\mu}_{i+1} \blacktriangleright \hbar_{2[ln]}$

$$\boldsymbol{\mu}_{i+1} = \boldsymbol{\mu}_i + \left\lfloor \frac{y_i (\boldsymbol{\mu}_i \cdot \boldsymbol{x}_i) - \epsilon}{0.5\alpha^{-1} - \boldsymbol{x}_i^\top \Sigma_i (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i)} \right\rfloor y_i \Sigma_i (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i)$$

where $\overline{\boldsymbol{x}}_i = \overline{x}_i \mathbf{1} \equiv \frac{\mathbf{1}^\top \Sigma_i \boldsymbol{x}_i}{\mathbf{1}^\top \Sigma_i \mathbf{1}} \mathbf{1}$

Lemma 25. $\alpha \blacktriangleright \hbar_{2[ln]}$, $\alpha = \left\lfloor \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \right\rfloor$ *such that*

$(a, b, c) = \left(\lambda' \left(\lambda' + \upsilon_i \phi^2\right), 2\lambda \left(\lambda' + \frac{\upsilon_i \phi^2}{2}\right), \lambda^2 - \upsilon_i \phi^2\right)$

$\left(\lambda, \lambda'\right) \approx \left((y_i (\boldsymbol{\mu}_i \cdot \boldsymbol{x}_i) - \epsilon)^2, 4\lambda \boldsymbol{x}_i^\top \Sigma_i (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i)\right)$

3.3.2. *Logarithm* : $\hbar_{1[\ln]}, \hbar_{2[\ln]}$.

$\Sigma_{i+1}^{-1} \blacktriangleright \hbar_{[1,2][\ln]}$, Lemma 16 $\equiv$ Lemma 26 $\equiv$ Lemma 6, $\Sigma_{i+1}^{-1} \blacktriangleright \hbar_{\emptyset[\ln]}$

$\Sigma_{i+1} \blacktriangleright \hbar_{[1,2][\ln]}$, Lemma 17 $\equiv$ Lemma 27 $\equiv$ Lemma 7, $\Sigma_{i+1} \blacktriangleright \hbar_{\emptyset[\ln]}$

$\sqrt{u_i} \blacktriangleright \hbar_{[1,2][\ln]}$, Lemma 18 $\equiv$ Lemma 28 $\equiv$ Lemma 8, $\sqrt{u_i} \blacktriangleright \hbar_{\emptyset[\ln]}$

Lemma 19. $\boldsymbol{\mu}_{i+1} \blacktriangleright \hbar_{1[\ln]}$

$$\boldsymbol{\mu}_{i+1} = \boldsymbol{\mu}_i + \langle y_i \ln (\boldsymbol{\mu}_i \cdot \boldsymbol{x}_i) - \epsilon \rangle \frac{\alpha y_i}{\boldsymbol{\mu}_i \cdot \boldsymbol{x}_i} \Sigma_i (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i)$$

where $\overline{\boldsymbol{x}}_i = \overline{x}_i \mathbf{1} \equiv \frac{\mathbf{1}^\top \Sigma_i \boldsymbol{x}_i}{\mathbf{1}^\top \Sigma_i \mathbf{1}} \mathbf{1}$

Lemma 20. $\alpha \blacktriangleright \hbar_{1[\ln]}$, $\alpha = \left\lfloor \frac{-b\pm\sqrt{b^2-4ac}}{2a} \right\rfloor$ *such that*

$(a,b,c) = \left(\lambda' \left(\lambda' + \upsilon_i \phi^2\right), 2\lambda \left(\lambda' + \frac{\upsilon_i \phi^2}{2}\right), \lambda^2 - \upsilon_i \phi^2\right)$

$\left(\lambda, \lambda'\right) \approx \left(y_i \ln\left(\boldsymbol{\mu}_i \cdot \boldsymbol{x}_i\right) - \epsilon, \frac{\boldsymbol{x}_i^\top \Sigma_i (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i)}{(\boldsymbol{\mu}_i \cdot \boldsymbol{x}_i)^2}\right)$

Lemma 29. $\boldsymbol{\mu}_{i+1} \blacktriangleright \hbar_{2[\ln]}$

$$\boldsymbol{\mu}_{i+1} \approx \boldsymbol{\mu}_i + \left\lfloor \frac{y_i \ln\left(\boldsymbol{\mu}_i \cdot \boldsymbol{x}_i\right) - \epsilon}{0.5\alpha^{-1} - \frac{\boldsymbol{x}_i^\top \Sigma_i (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i)}{(\boldsymbol{\mu}_i \cdot \boldsymbol{x}_i)^2}} \right\rfloor \frac{y_i}{\boldsymbol{\mu}_i \cdot \boldsymbol{x}_i} \Sigma_i \left(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i\right)$$

where $\overline{\boldsymbol{x}}_i = \overline{x}_i \mathbf{1} \equiv \frac{\mathbf{1}^\top \Sigma_i \boldsymbol{x}_i}{\mathbf{1}^\top \Sigma_i \mathbf{1}} \mathbf{1}$

Lemma 30. $\alpha \blacktriangleright \hbar_{2[\ln]}$, $\alpha = \left\lfloor \frac{-b\pm\sqrt{b^2-4ac}}{2a} \right\rfloor$ *such that*

$(a,b,c) = \left(\lambda' \left(\lambda' + \upsilon_i \phi^2\right), 2\lambda \left(\lambda' + \frac{\upsilon_i \phi^2}{2}\right), \lambda^2 - \upsilon_i \phi^2\right)$

$\left(\lambda, \lambda'\right) \approx \left(\left(y_i \ln\left(\boldsymbol{\mu}_i \cdot \boldsymbol{x}_i\right) - \epsilon\right)^2, 4\lambda \frac{\boldsymbol{x}_i^\top \Sigma_i (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i)}{(\boldsymbol{\mu}_i \cdot \boldsymbol{x}_i)^2}\right)$

4. Implementation

4.1. Results in section 3. is valid for the $\overset{\leftrightarrow}{\triangle}$ simplex. A more common constraint is $\triangle$ simplex; however the close-form solution is not possible with this simplex. Projecting simplex $\overset{\leftrightarrow}{\triangle}$ on $\triangle$ is a practical approximation; [LHZG11] reports the effectiveness of this method. The projection necessarily requires a certain transformation of Σ-covariance matrix. Further information on implementing projection algorithm and covariance transformation is in [CY11] and [LHZG11], respectively.

`Conjecture. Correlation transform is an nSD-effective covariance transformer.`

4.2. Section 3 presents various choices of simplex, from which one can limit the set of simplex using statistical dominance concept, e.g. *nSD-effective*. Then projecting the simplex and integrating or *fusing* them which is, in practice, an empirical issue. We define a new simplex fusing method *FED (fusing extensive dimension)* as follows. Let $\triangle_{i\in\{1\ldots m\}}$ be a set of *nSD-effective* simplex, each $\triangle_i \in [0,1]^N$. Connect m subsimplex into a vector in $[0,1]^{m\cdot N}$; apply simplex projection to the vector. The result is simplex $\triangle \in [0,1]^{m\cdot N}$; overlay simplex $\triangle$, i.e. slot $\triangle$ into m vectors in $[0,1]^N$ and sum the vectors with the proper array. The overlay will compose a *FED* simplex $\in [0,1]^N$.

`Conjecture. FED simplex is an nSD-effective fuse of its nSD-effective subsimplex.`

‡ *nSD-effective* is empirical non-dominated, wrt. to the n-order stochastic dominance definition [Dav06].

5. Remark

5.1. The logic of confidence constraint. Suppose $\frac{F(\boldsymbol{w}\cdot\boldsymbol{x_i}) - \mu_{F(\boldsymbol{w}\cdot\boldsymbol{x_i})}}{\sigma_{F(\boldsymbol{w}\cdot\boldsymbol{x_i})}} = Z_{\Phi-\mathtt{cdf}}$; consider a generic confidence constraint $\Pr\left(F\left(\boldsymbol{w}\cdot\boldsymbol{x_i}\right) \geq 0\right) \geq \eta \equiv \Phi\left(\phi\right)$.

$$\Pr\left(\frac{F\left(\boldsymbol{w}\cdot\boldsymbol{x_i}\right) - \mu_{F(\boldsymbol{w}\cdot\boldsymbol{x_i})}}{\sigma_{F(\boldsymbol{w}\cdot\boldsymbol{x_i})}} \geq \frac{-\mu_{F(\boldsymbol{w}\cdot\boldsymbol{x_i})}}{\sigma_{F(\boldsymbol{w}\cdot\boldsymbol{x_i})}}\right) \geq \eta \Rightarrow \Phi\left(\frac{-\mu_{F(\boldsymbol{w}\cdot\boldsymbol{x_i})}}{\sigma_{F(\boldsymbol{w}\cdot\boldsymbol{x_i})}}\right) \leq 1-\eta$$

$$\frac{-\mu_{F(\boldsymbol{w}\cdot\boldsymbol{x_i})}}{\sigma_{F(\boldsymbol{w}\cdot\boldsymbol{x_i})}} \leq \Phi^{-1}\left(1-\eta\right) = -\Phi^{-1}\left(\eta\right) \Rightarrow \mu_{F(\boldsymbol{w}\cdot\boldsymbol{x_i})} \geq \Phi^{-1}\left(\eta\right)\sigma_{F(\boldsymbol{w}\cdot\boldsymbol{x_i})} = \phi\sigma_{F(\boldsymbol{w}\cdot\boldsymbol{x_i})}$$

, i.e. the distance $\left|\mu_{F(\boldsymbol{w}\cdot\boldsymbol{x_i})} - F\left(\mu_{\boldsymbol{w}\cdot\boldsymbol{x_i}}\right) - \phi\left(\sigma_{F(\boldsymbol{w}\cdot\boldsymbol{x_i})} - \sigma_{\boldsymbol{w}\cdot\boldsymbol{x_i}}\right)\right|$ determines the proximity to the confidence constraint; [OC09] discusses the validity of similar approach for online optimization.

5.2. The approximating property of $\left[\text{normal,hinge, hinge}^2\right]$ confidence. Define $\left[\text{normal,hinge, hinge}^2\right]$ function as follows,

normal: $\hbar_{\emptyset[f]} \in \left\{\hbar_{\emptyset[ln]}, \hbar_{\emptyset[\ln]}\right\} \equiv \left\{y_i\left(\boldsymbol{\mu}\cdot\boldsymbol{x}_i\right)-\epsilon, y_i\ln\left(\boldsymbol{\mu}\cdot\boldsymbol{x}_i\right)-\epsilon\right\}$
hinge: $\hbar_{1[f]} \in \left\{\hbar_{1[ln]}, \hbar_{1[\ln]}\right\} \equiv \left\{\lfloor y_i\left(\boldsymbol{\mu}\cdot\boldsymbol{x}_i\right)-\epsilon\rfloor, \lfloor y_i\ln\left(\boldsymbol{\mu}\cdot\boldsymbol{x}_i\right)-\epsilon\rfloor\right\}$
hinge^2: $\hbar_{2[f]} \in \left\{\hbar_{2[ln]}, \hbar_{2[\ln]}\right\} \equiv \left\{\lfloor y_i\left(\boldsymbol{\mu}\cdot\boldsymbol{x}_i\right)-\epsilon\rfloor^2, \lfloor y_i\ln\left(\boldsymbol{\mu}\cdot\boldsymbol{x}_i\right)-\epsilon\rfloor^2\right\}$

, as a result of assumption $\boldsymbol{w} \sim \mathbb{N}\left(\boldsymbol{\mu}, \Sigma = \Upsilon^2\right)$;

normal: $\hbar_{\emptyset[ln]}$ is exact; $\hbar_{\emptyset[\ln]}$ is approximate

$F\left(\boldsymbol{w}\cdot\boldsymbol{x_i}\right) = y_i\left(\boldsymbol{w}\cdot\boldsymbol{x}_i\right)-\epsilon \Rightarrow \left(\mu_{F(\boldsymbol{w}\cdot\boldsymbol{x_i})}, \sigma^2_{F(\boldsymbol{w}\cdot\boldsymbol{x_i})}\right) = \left(y_i\left(\boldsymbol{\mu}\cdot\boldsymbol{x}_i\right)-\epsilon, \sigma_{\boldsymbol{w}\cdot\boldsymbol{x}_i} = \boldsymbol{x}_i^\top\Sigma\boldsymbol{x}_i\right)$
$F\left(\boldsymbol{w}\cdot\boldsymbol{x_i}\right) = y_i\ln\left(\boldsymbol{w}\cdot\boldsymbol{x}_i\right)-\epsilon \Rightarrow \mu_{F(\boldsymbol{w}\cdot\boldsymbol{x_i})} \approx y_i\ln\left(\boldsymbol{\mu}\cdot\boldsymbol{x}_i\right)$

hinge: $\hbar_{1[ln][\ln]}$ is approximate

$F\left(\boldsymbol{w}\cdot\boldsymbol{x_i}\right) = \lfloor y_i\left(\boldsymbol{w}\cdot\boldsymbol{x}_i\right)-\epsilon\rfloor \Rightarrow \mu_{F(\boldsymbol{w}\cdot\boldsymbol{x_i})} \approx \lfloor y_i\left(\boldsymbol{\mu}\cdot\boldsymbol{x}_i\right)-\epsilon\rfloor$
$F\left(\boldsymbol{w}\cdot\boldsymbol{x_i}\right) = \lfloor y_i\ln\left(\boldsymbol{w}\cdot\boldsymbol{x}_i\right)-\epsilon\rfloor \Rightarrow \mu_{F(\boldsymbol{w}\cdot\boldsymbol{x_i})} \approx \lfloor y_i\ln\left(\boldsymbol{\mu}\cdot\boldsymbol{x}_i\right)-\epsilon\rfloor$

hinge^2: $\hbar_{2[ln][\ln]}$ is approximate

$F\left(\boldsymbol{w}\cdot\boldsymbol{x_i}\right) = \lfloor y_i\left(\boldsymbol{w}\cdot\boldsymbol{x}_i\right)-\epsilon\rfloor^2 \Rightarrow \mu_{F(\boldsymbol{w}\cdot\boldsymbol{x_i})} \approx \lfloor y_i\left(\boldsymbol{\mu}\cdot\boldsymbol{x}_i\right)-\epsilon\rfloor^2$
$F\left(\boldsymbol{w}\cdot\boldsymbol{x_i}\right) = \lfloor y_i\ln\left(\boldsymbol{w}\cdot\boldsymbol{x}_i\right)-\epsilon\rfloor^2 \Rightarrow \mu_{F(\boldsymbol{w}\cdot\boldsymbol{x_i})} \approx \lfloor y_i\ln\left(\boldsymbol{\mu}\cdot\boldsymbol{x}_i\right)-\epsilon\rfloor^2$

Appendix

Lemma 1. $\Sigma^{-1}_{i+1} \blacktriangleright \hbar_{\emptyset[ln]}$

$$\frac{\partial}{\partial\Upsilon}\mathcal{L} = 0 = -\Upsilon^{-1} + \frac{1}{2}\Upsilon_i^{-2}\Upsilon + \frac{1}{2}\Upsilon\Upsilon_i^{-2} + \alpha\phi\frac{\boldsymbol{x}_i\boldsymbol{x}_i^\top\Upsilon}{2\sqrt{\boldsymbol{x}_i^\top\Upsilon^2\boldsymbol{x}_i}} + \alpha\phi\frac{\Upsilon\boldsymbol{x}_i\boldsymbol{x}_i^\top}{2\sqrt{\boldsymbol{x}_i^\top\Upsilon^2\boldsymbol{x}_i}}$$

Υ^{-1} update condition is,

$$\Upsilon^{-1} = \frac{1}{2}\Upsilon_i^{-2}\Upsilon + \frac{1}{2}\Upsilon\Upsilon_i^{-2} + \alpha\phi\frac{\boldsymbol{x}_i\boldsymbol{x}_i^\top\Upsilon}{2\sqrt{\boldsymbol{x}_i^\top\Upsilon^2\boldsymbol{x}_i}} + \alpha\phi\frac{\Upsilon\boldsymbol{x}_i\boldsymbol{x}_i^\top}{2\sqrt{\boldsymbol{x}_i^\top\Upsilon^2\boldsymbol{x}_i}} \qquad \left[\Upsilon^{-1}\right]$$

Start with the solution, Υ^{-2} implicit update,

$$\Upsilon^{-2} \equiv \Upsilon^{-2}_{i+1} = \Upsilon_i^{-2} + \alpha\phi\frac{\boldsymbol{x}_i\boldsymbol{x}_i^\top}{\sqrt{\boldsymbol{x}_i^\top\Upsilon^2\boldsymbol{x}_i}} \qquad \left[\Upsilon^{-2}\right]$$

which yields

$$\frac{\Upsilon^{-1}}{2} = \frac{\Upsilon_i^{-2}\Upsilon}{2} + \frac{\alpha\phi}{2} \cdot \frac{\boldsymbol{x}_i \boldsymbol{x}_i^\top \Upsilon}{\sqrt{\boldsymbol{x}_i^\top \Upsilon^2 \boldsymbol{x}_i}} \qquad [\times\Upsilon]$$

$$\frac{\Upsilon^{-1}}{2} = \frac{\Upsilon\Upsilon_i^{-2}}{2} + \frac{\alpha\phi}{2} \cdot \frac{\Upsilon \boldsymbol{x}_i \boldsymbol{x}_i^\top}{\sqrt{\boldsymbol{x}_i^\top \Upsilon^2 \boldsymbol{x}_i}} \qquad [\Upsilon\times]$$

$\left[\Upsilon^{-2}\right] \Rightarrow [\times\Upsilon] + [\Upsilon\times] \Rightarrow \left[\Upsilon^{-1}\right]$, i.e. $\Upsilon^{-2}-$implicit update satisfying $\Upsilon^{-1}-$update. The result is direct from the replacement $\left(\Upsilon_i^2, \Upsilon^2\right) = (\Sigma_i, \Sigma_{i+1})$:

$$\Sigma_{i+1}^{-1} = \Sigma_i^{-1} + \alpha\phi \frac{\boldsymbol{x}_i \boldsymbol{x}_i^\top}{\sqrt{\boldsymbol{x}_i^\top \Sigma_{i+1} \boldsymbol{x}_i}}$$

□

Lemma 2. $\Sigma_{i+1} \blacktriangleright \hbar_{\emptyset[ln]}$

Apply matrix inversion to $\Sigma_{i+1}^{-1} = \Sigma_i^{-1} + \alpha\phi \frac{\boldsymbol{x}_i \boldsymbol{x}_i^\top}{\sqrt{\boldsymbol{x}_i^\top \Sigma_{i+1} \boldsymbol{x}_i}}$,

$$\Sigma_{i+1} = \Sigma_i - \frac{\Sigma_i \boldsymbol{x}_i \boldsymbol{x}_i^\top \Sigma_i}{\frac{\sqrt{\boldsymbol{x}_i^\top \Sigma_{i+1} \boldsymbol{x}_i}}{\alpha\phi} + \boldsymbol{x}_i^\top \Sigma_i \boldsymbol{x}_i} = \Sigma_i - \frac{\alpha\phi \Sigma_i \boldsymbol{x}_i \boldsymbol{x}_i^\top \Sigma_i}{\sqrt{\boldsymbol{x}_i^\top \Sigma_{i+1} \boldsymbol{x}_i} + \alpha\phi \boldsymbol{x}_i^\top \Sigma_i \boldsymbol{x}_i}$$

$$\Sigma_{i+1} = \Sigma_i - \frac{\alpha\phi \Sigma_i \boldsymbol{x}_i \boldsymbol{x}_i^\top \Sigma_i}{\sqrt{u_i} + \alpha\phi \upsilon_i} = \Sigma_i - \beta \Sigma_i \boldsymbol{x}_i \boldsymbol{x}_i^\top \Sigma_i$$

□

Lemma 3. $\sqrt{u_i} \blacktriangleright \hbar_{\emptyset[ln]}$

$$\Sigma_{i+1} = \Sigma_i - \frac{\alpha\phi \Sigma_i \boldsymbol{x}_i \boldsymbol{x}_i^\top \Sigma_i}{\sqrt{u_i} + \alpha\phi \upsilon_i} \Rightarrow \boldsymbol{x}_i^\top \Sigma_{i+1} \boldsymbol{x}_i = \boldsymbol{x}_i^\top \Sigma_i \boldsymbol{x}_i - \frac{\alpha\phi \left(\boldsymbol{x}_i^\top \Sigma_i \boldsymbol{x}_i\right) \left(\boldsymbol{x}_i^\top \Sigma_i \boldsymbol{x}_i\right)}{\sqrt{u_i} + \alpha\phi \upsilon_i}$$

$$u_i = \upsilon_i - \frac{\alpha\phi \upsilon_i^2}{\sqrt{u_i} + \alpha\phi \upsilon_i} \Rightarrow \sqrt{u_i} = \frac{-\alpha\phi\upsilon_i + \sqrt{\alpha^2\phi^2\upsilon_i^2 + 4\upsilon_i}}{2}$$

□

Lemma 4. $\boldsymbol{\mu}_{i+1} \blacktriangleright \hbar_{\emptyset[ln]}$

$$\frac{\partial}{\partial \boldsymbol{\mu}} \mathcal{L} = 0 = \Upsilon_i^{-2} \left(\boldsymbol{\mu} - \boldsymbol{\mu}_i\right) - \alpha \hbar_\emptyset^{'} f^{'} y_i \boldsymbol{x}_i + \rho \mathbf{1} \,;\ \frac{\partial}{\partial \rho} \mathcal{L} = 0 = \boldsymbol{\mu} \cdot \mathbf{1} - 1$$

$$\Upsilon_i^{-2} \left(\boldsymbol{\mu} - \boldsymbol{\mu}_i\right) - \alpha \hbar_\emptyset^{'} f^{'} y_i \boldsymbol{x}_i + \rho\ \mathbf{1} = 0 \Rightarrow \boldsymbol{\mu} = \boldsymbol{\mu}_i + \Upsilon_i^2 \left(\alpha \hbar_\emptyset^{'} f^{'} y_i \boldsymbol{x}_i - \rho \mathbf{1}\right)$$

$$\mathbf{1}^\top \boldsymbol{\mu} = \mathbf{1}^\top \boldsymbol{\mu}_i + \alpha \hbar_\emptyset^{'} f^{'} y_i \mathbf{1}^\top \Upsilon_i^2 \boldsymbol{x}_i - \rho \mathbf{1}^\top \Upsilon_i^2 \mathbf{1}$$

$$\rho \mathbf{1} = \alpha \hbar_\emptyset^{'} f^{'} y_i \left(\frac{\mathbf{1}^\top \Upsilon_i^2 \boldsymbol{x}_i}{\mathbf{1}^\top \Upsilon_i^2 \mathbf{1}}\right) \mathbf{1} = \alpha \hbar_\emptyset^{'} f^{'} y_i \overline{\boldsymbol{x}}_i \Rightarrow \boldsymbol{\mu} = \boldsymbol{\mu}_i + \alpha \hbar_\emptyset^{'} f^{'} y_i \Upsilon_i^2 \left(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i\right)$$

use $\hbar_\emptyset^{'}(.) = 1$, $f^{'}(.) = 1$ and $\Upsilon_i^2 = \Sigma_i$ to have $\boldsymbol{\mu}_{i+1} = \boldsymbol{\mu} = \boldsymbol{\mu}_i + \alpha y_i \Sigma_i \left(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i\right)$ □

Lemma 5. $\alpha \blacktriangleright \hbar_{\emptyset[ln]}$

From Lemma 3 $\sqrt{u_i} = \frac{-\alpha\phi\upsilon_i + \sqrt{\alpha^2\phi^2\upsilon_i^2 + 4\upsilon_i}}{2}$, which can be simplified with $\lambda + \lambda^{'}\alpha = \phi\sqrt{u_i}$. Its quadratic is $a\alpha^2 + b\alpha + c = 0$, such that $(a, b, c) = \left(\lambda^{'}\left(\lambda^{'} + \upsilon_i\phi^2\right), 2\lambda\left(\lambda^{'} + \frac{\upsilon_i\phi^2}{2}\right), \lambda^2 - \upsilon_i\phi^2\right)$. The solution to $\lambda + \lambda^{'}\alpha = \phi\sqrt{u_i}$ is $\frac{-b\pm\sqrt{b^2-4ac}}{2a}$. We choose $\alpha = \left\lfloor \frac{-b\pm\sqrt{b^2-4ac}}{2a} \right\rfloor$ to ensure valid $\alpha \geq 0$.

To find $\left(\lambda, \lambda^{'}\right)$, use binding constraint $\phi \left\|\Upsilon \boldsymbol{x}_i\right\| = \hbar_{\emptyset[ln]} \Rightarrow \phi \left\|\Upsilon \boldsymbol{x}_i\right\| = y_i\left(\boldsymbol{\mu}\cdot\boldsymbol{x}_i\right) - \epsilon$. Apply the update $\boldsymbol{\mu} = \boldsymbol{\mu}_i + \alpha y_i \Sigma_i\left(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i\right)$ and $\sqrt{u_i} \equiv \left\|\Upsilon \boldsymbol{x}_i\right\|$,

$$\phi\sqrt{u_i} = y_i\boldsymbol{\mu}_i\cdot\boldsymbol{x}_i - \epsilon + \alpha\boldsymbol{x}_i^{\top}\Sigma_i\left(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i\right)$$

i.e. $\left(\lambda, \lambda^{'}\right) = \left(y_i\boldsymbol{\mu}_i\cdot\boldsymbol{x}_i - \epsilon, \boldsymbol{x}_i^{\top}\Sigma_i\left(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i\right)\right)$. ☐

Lemma 6. $\Sigma_{i+1}^{-1} \blacktriangleright \hbar_{\emptyset[\ln]}$

≡ Lemma 1. ☐

Lemma 7. $\Sigma_{i+1} \blacktriangleright \hbar_{\emptyset[\ln]}$

≡ Lemma 2. ☐

Lemma 8. $\sqrt{u_i} \blacktriangleright \hbar_{\emptyset[\ln]}$

≡ Lemma 3. ☐

Lemma 9. $\boldsymbol{\mu}_{i+1} \blacktriangleright \hbar_{\emptyset[\ln]}$

Similar to Lemma 4, $\boldsymbol{\mu} = \boldsymbol{\mu}_i + \alpha\hbar_{\emptyset}^{'}f^{'}y_i\Upsilon_i^2\left(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i\right)$; use $\hbar_{\emptyset}^{'}(.) = 1$; $\ln\left(\boldsymbol{\mu}\cdot\boldsymbol{x}_i\right) \approx \ln\left(\boldsymbol{\mu}_i\cdot\boldsymbol{x}_i\right) + \frac{\left(\boldsymbol{\mu}-\boldsymbol{\mu}_i\right)\cdot\boldsymbol{x}_i}{\boldsymbol{\mu}_i\cdot\boldsymbol{x}_i} \Rightarrow f^{'}(.) = \frac{1}{\boldsymbol{\mu}_i\cdot\boldsymbol{x}_i}$ and $\Upsilon_i^2 = \Sigma_i$, which gives $\boldsymbol{\mu}_{i+1} = \boldsymbol{\mu} \approx \boldsymbol{\mu}_i + \frac{\alpha y_i}{\boldsymbol{\mu}_i\cdot\boldsymbol{x}_i}\Sigma_i\left(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i\right)$ ☐

Lemma 10. $\alpha \blacktriangleright \hbar_{\emptyset[\ln]}$

Similar to Lemma 5, $\alpha = \left\lfloor \frac{-b\pm\sqrt{b^2-4ac}}{2a} \right\rfloor$ where $(a, b, c) = \left(\lambda^{'}\left(\lambda^{'} + \upsilon_i\phi^2\right), 2\lambda\left(\lambda^{'} + \frac{\upsilon_i\phi^2}{2}\right), \lambda^2 - \upsilon_i\phi^2\right)$.

To find $\left(\lambda, \lambda^{'}\right)$, set the constraint binding $\phi \left\|\Upsilon \boldsymbol{x}_i\right\| = \hbar_{\emptyset[\ln]} \Rightarrow \phi \left\|\Upsilon \boldsymbol{x}_i\right\| = y_i\ln\left(\boldsymbol{\mu}\cdot\boldsymbol{x}_i\right) - \epsilon$. Apply the update $\boldsymbol{\mu} = \boldsymbol{\mu}_i + \frac{\alpha y_i}{\boldsymbol{\mu}_i\cdot\boldsymbol{x}_i}\Sigma_i\left(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i\right)$ and $\sqrt{u_i} \equiv \left\|\Upsilon \boldsymbol{x}_i\right\|$ and the approximation $y_i\ln\left(\boldsymbol{\mu}\cdot\boldsymbol{x}_i\right) - \epsilon \approx y_i\left(\ln\left(\boldsymbol{\mu}_i\cdot\boldsymbol{x}_i\right) + \frac{\left(\boldsymbol{\mu}-\boldsymbol{\mu}_i\right)\cdot\boldsymbol{x}_i}{\boldsymbol{\mu}_i\cdot\boldsymbol{x}_i}\right) - \epsilon$

$$\phi\sqrt{u_i} \approx y_i\ln\left(\boldsymbol{\mu}_i\cdot\boldsymbol{x}_i\right) - \epsilon + \alpha\frac{\boldsymbol{x}_i^{\top}\Sigma_i\left(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i\right)}{\left(\boldsymbol{\mu}_i\cdot\boldsymbol{x}_i\right)^2}.$$

, i.e. $\left(\lambda, \lambda^{'}\right) \approx \left(y_i\ln\left(\boldsymbol{\mu}_i\cdot\boldsymbol{x}_i\right) - \epsilon, \frac{\boldsymbol{x}_i^{\top}\Sigma_i\left(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i\right)}{\left(\boldsymbol{\mu}_i\cdot\boldsymbol{x}_i\right)^2}\right)$. ☐

Lemma 11. $\Sigma_{i+1}^{-1} \blacktriangleright \hbar_{1[ln]}$

≡ Lemma 1. ☐

Lemma 12. $\Sigma_{i+1} \blacktriangleright \hbar_{1[ln]}$

≡ Lemma 2. ☐

Lemma 13. $\sqrt{u_i} \blacktriangleright \hbar_{1[ln]}$

$\equiv$ Lemma 3. $\square$

Lemma 14. $\boldsymbol{\mu}_{i+1} \blacktriangleright \hbar_{1[ln]}$

Similar to Lemma Lemma 4, $\boldsymbol{\mu} = \boldsymbol{\mu}_i + \alpha \hbar_1^{'} f^{'} y_i \Upsilon_i^2 (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i)$. There are two cases, $y_i (\boldsymbol{\mu} \cdot \boldsymbol{x}_i) - \epsilon [>] [\leq] 0$.

Case $[>]$: $\hbar_1^{'} (.) = 1$, $f^{'} (.) = 1$ and $\Upsilon_i^2 = \Sigma_i$, $\Rightarrow \boldsymbol{\mu}_{i+1} = \boldsymbol{\mu} = \boldsymbol{\mu}_i + \alpha y_i \Sigma_i (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i)$
Case $[\leq]$: $\hbar_1^{'} (.) = 0 \Rightarrow \boldsymbol{\mu}_{i+1} = \boldsymbol{\mu} = \boldsymbol{\mu}_i$
With some manipulation we find a $\boldsymbol{\mu}-$update

$$\boldsymbol{\mu}_{i+1} = \boldsymbol{\mu}_i + \langle y_i (\boldsymbol{\mu}_i \cdot \boldsymbol{x}_i) - \epsilon \rangle \alpha y_i \Sigma_i (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i)$$

$\square$

Lemma 15. $\alpha \blacktriangleright \hbar_{1[ln]}$

Similar to Lemma 5, $\alpha = \left\lfloor \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \right\rfloor$ where $(a, b, c) = \left(\lambda^{'} \left(\lambda^{'} + \upsilon_i \phi^2\right), 2\lambda \left(\lambda^{'} + \frac{\upsilon_i \phi^2}{2}\right), \lambda^2 - \upsilon_i \phi^2\right)$.
To find $\left(\lambda, \lambda^{'}\right)$, use binding constraint $\phi \|\Upsilon \boldsymbol{x}_i\| = \hbar_{1[ln]} \Rightarrow \phi \|\Upsilon \boldsymbol{x}_i\| = \lfloor y_i (\boldsymbol{\mu} \cdot \boldsymbol{x}_i) - \epsilon \rfloor$. We only need the update-case $y_i (\boldsymbol{\mu} \cdot \boldsymbol{x}_i) - \epsilon > 0$. Apply the update $\boldsymbol{\mu} = \boldsymbol{\mu}_i + \alpha y_i \Sigma_i (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i)$ and $\sqrt{u_i} \equiv \|\Upsilon \boldsymbol{x}_i\|$,

$$\phi \sqrt{u_i} = \lfloor y_i (\boldsymbol{\mu} \cdot \boldsymbol{x}_i) - \epsilon \rfloor = y_i (\boldsymbol{\mu} \cdot \boldsymbol{x}_i) - \epsilon = y_i \boldsymbol{\mu}_i \cdot \boldsymbol{x}_i - \epsilon + \alpha \boldsymbol{x}_i^\top \Sigma_i (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i)$$

i.e. $\left(\lambda, \lambda^{'}\right) = \left(y_i \boldsymbol{\mu}_i \cdot \boldsymbol{x}_i - \epsilon, \boldsymbol{x}_i^\top \Sigma_i (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i)\right)$. $\square$

Lemma 16. $\Sigma_{i+1}^{-1} \blacktriangleright \hbar_{1[\text{ln}]}$

$\equiv$ Lemma 6. $\square$

Lemma 17. $\Sigma_{i+1} \blacktriangleright \hbar_{1[\text{ln}]}$

$\equiv$ Lemma 7. $\square$

Lemma 18. $\sqrt{u_i} \blacktriangleright \hbar_{1[\text{ln}]}$

$\equiv$ Lemma 8. $\square$

Lemma 19. $\boldsymbol{\mu}_{i+1} \blacktriangleright \hbar_{1[\text{ln}]}$

Similar to Lemma 14, $\boldsymbol{\mu} = \boldsymbol{\mu}_i + \alpha \hbar_1^{'} f^{'} y_i \Upsilon_i^2 (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i)$ with two cases, $y_i \ln (\boldsymbol{\mu} \cdot \boldsymbol{x}_i) - \epsilon [>] [\leq] 0$.

Case $[>]$: $\hbar_1^{'} (.) = 1$, $\ln (\boldsymbol{\mu} \cdot \boldsymbol{x}_i) \approx \ln (\boldsymbol{\mu}_i \cdot \boldsymbol{x}_i) + \frac{(\boldsymbol{\mu} - \boldsymbol{\mu}_i) \cdot \boldsymbol{x}_i}{\boldsymbol{\mu}_i \cdot \boldsymbol{x}_i} \Rightarrow f^{'} (.) = \frac{1}{\boldsymbol{\mu}_i \cdot \boldsymbol{x}_i}$ and $\Upsilon_i^2 = \Sigma_i$

$$\boldsymbol{\mu}_{i+1} = \boldsymbol{\mu} = \boldsymbol{\mu}_i + \frac{\alpha y_i}{\boldsymbol{\mu}_i \cdot \boldsymbol{x}_i} \Sigma_i (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i)$$

Case $[\leq]$: $\hbar_1^{'} (.) = 0 \Rightarrow \boldsymbol{\mu}_{i+1} = \boldsymbol{\mu} = \boldsymbol{\mu}_i$
With some manipulation we find a $\boldsymbol{\mu}-$update

$$\boldsymbol{\mu}_{i+1} = \boldsymbol{\mu}_i + \langle y_i \ln (\boldsymbol{\mu}_i \cdot \boldsymbol{x}_i) - \epsilon \rangle \frac{\alpha y_i}{\boldsymbol{\mu}_i \cdot \boldsymbol{x}_i} \Sigma_i (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i)$$

$\square$

Lemma 20. $\alpha \blacktriangleright \hbar_{1[\ln]}$

Similar to Lemma 15, $\alpha = \left\lfloor \frac{-b\pm\sqrt{b^2-4ac}}{2a} \right\rfloor$ where $(a,b,c) = \left(\lambda^{'}\left(\lambda^{'} + \upsilon_i\phi^2\right), 2\lambda\left(\lambda^{'} + \frac{\upsilon_i\phi^2}{2}\right), \lambda^2 - \upsilon_i\phi^2\right)$.

To find $\left(\lambda, \lambda^{'}\right)$, set the constraint binding $\phi\left\|\Upsilon \boldsymbol{x}_i\right\| = \hbar_{1[\ln]} \Rightarrow \phi\left\|\Upsilon \boldsymbol{x}_i\right\| = \lfloor y_i \ln\left(\boldsymbol{\mu}\cdot\boldsymbol{x}_i\right) - \epsilon\rfloor$. We only need the update-case $y_i\ln\left(\boldsymbol{\mu}\cdot\boldsymbol{x}_i\right) - \epsilon > 0$. Apply the update $\boldsymbol{\mu} = \boldsymbol{\mu}_i + \frac{\alpha y_i}{\boldsymbol{\mu}_i\cdot\boldsymbol{x}_i}\Sigma_i\left(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i\right)$ and $\sqrt{u_i} \equiv \left\|\Upsilon\boldsymbol{x}_i\right\|$ and the approximation $y_i\ln\left(\boldsymbol{\mu}\cdot\boldsymbol{x}_i\right) - \epsilon \approx y_i\left(\ln\left(\boldsymbol{\mu}_i\cdot\boldsymbol{x}_i\right) + \frac{(\boldsymbol{\mu}-\boldsymbol{\mu}_i)\cdot\boldsymbol{x}_i}{\boldsymbol{\mu}_i\cdot\boldsymbol{x}_i}\right) - \epsilon$,

$$\phi\sqrt{u_i} = \lfloor y_i\ln\left(\boldsymbol{\mu}\cdot\boldsymbol{x}_i\right) - \epsilon\rfloor = y_i\ln\left(\boldsymbol{\mu}\cdot\boldsymbol{x}_i\right) - \epsilon \approx y_i\ln\left(\boldsymbol{\mu}_i\cdot\boldsymbol{x}_i\right) - \epsilon + \alpha\frac{\boldsymbol{x}_i^\top\Sigma_i\left(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i\right)}{\left(\boldsymbol{\mu}_i\cdot\boldsymbol{x}_i\right)^2}$$

, i.e. $\left(\lambda, \lambda^{'}\right) \approx \left(y_i\ln\left(\boldsymbol{\mu}_i\cdot\boldsymbol{x}_i\right) - \epsilon, \frac{\boldsymbol{x}_i^\top\Sigma_i\left(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i\right)}{\left(\boldsymbol{\mu}_i\cdot\boldsymbol{x}_i\right)^2}\right)$. □

Lemma 21. $\Sigma_{i+1}^{-1} \blacktriangleright \hbar_{2[ln]}$

≡ Lemma 1. □

Lemma 22. $\Sigma_{i+1} \blacktriangleright \hbar_{2[ln]}$

≡ Lemma 2. □

Lemma 23. $\sqrt{u_i} \blacktriangleright \hbar_{2[ln]}$

≡ Lemma 3. □

Lemma 24. $\boldsymbol{\mu}_{i+1} \blacktriangleright \hbar_{2[ln]}$

Similar to Lemma Lemma 4, $\boldsymbol{\mu} = \boldsymbol{\mu}_i + \alpha\hbar_2^{'}f^{'}y_i\Upsilon_i^2\left(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i\right)$. There are two cases, $y_i\left(\boldsymbol{\mu}\cdot\boldsymbol{x}_i\right) - \epsilon\,[>]\,[\leq]\,0$.

Case $[>]$: $\hbar_2^{'}(.) = 2\left(y_i\left(\boldsymbol{\mu}\cdot\boldsymbol{x}_i\right) - \epsilon\right)$; use $f^{'}(.) = 1$ and $\Upsilon_i^2 = \Sigma_i$,

$$\boldsymbol{\mu} = \boldsymbol{\mu}_i + 2\alpha\left(y_i\left(\boldsymbol{\mu}\cdot\boldsymbol{x}_i\right) - \epsilon\right)y_i\Sigma_i\left(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i\right)$$

$$y_i\left(\boldsymbol{\mu}\cdot\boldsymbol{x}_i\right) - \epsilon = y_i\left(\boldsymbol{\mu}_i\cdot\boldsymbol{x}_i\right) - \epsilon + 2\alpha\left(y_i\left(\boldsymbol{\mu}\cdot\boldsymbol{x}_i\right) - \epsilon\right)\boldsymbol{x}_i^\top\Sigma_i\left(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i\right)$$

Write $X = y_i\left(\boldsymbol{\mu}\cdot\boldsymbol{x}_i\right) - \epsilon, C = y_i\left(\boldsymbol{\mu}_i\cdot\boldsymbol{x}_i\right) - \epsilon, S = 2\alpha\boldsymbol{x}_i^\top\Sigma_i\left(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i\right)$,

$$\left(\boldsymbol{\mu}, X\right) = \left(\boldsymbol{\mu}_i + 2\alpha X y_i\Sigma_i\left(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i\right), C + SX = \frac{C}{1-S}\right)$$

Case $[\leq]$: $\hbar_2^{'}(.) = 0 \Rightarrow \left(\boldsymbol{\mu}, X\right) = \left(\boldsymbol{\mu}_i + 2\alpha X y_i\Sigma_i\left(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i\right), 0\right)$.

We can conclude the update $\boldsymbol{\mu}_{i+1} = \boldsymbol{\mu} = \boldsymbol{\mu}_i + 2\alpha\lfloor X\rfloor y_i\Sigma_i\left(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i\right)$

$$\boldsymbol{\mu}_{i+1} = \boldsymbol{\mu}_i + \left\lfloor\frac{y_i\left(\boldsymbol{\mu}_i\cdot\boldsymbol{x}_i\right) - \epsilon}{0.5\alpha^{-1} - \boldsymbol{x}_i^\top\Sigma_i\left(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i\right)}\right\rfloor y_i\Sigma_i\left(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i\right)$$

□

Lemma 25. $\alpha \blacktriangleright \hbar_{2[ln]}$

Similar to Lemma 5, $\alpha = \left\lfloor \frac{-b \pm \sqrt{b^2-4ac}}{2a} \right\rfloor$ where $(a,b,c) = \left(\lambda^{'}\left(\lambda^{'} + \upsilon_i \phi^2\right), 2\lambda\left(\lambda^{'} + \frac{\upsilon_i \phi^2}{2}\right), \lambda^2 - \upsilon_i \phi^2\right)$.

To find $\left(\lambda, \lambda^{'}\right)$, use binding constraint $0 \leq \phi \left\|\Upsilon \boldsymbol{x}_i\right\| = \hbar_{2[ln]} \Rightarrow \phi \left\|\Upsilon \boldsymbol{x}_i\right\| = \left\lfloor y_i \left(\boldsymbol{\mu} \cdot \boldsymbol{x}_i\right) - \epsilon \right\rceil^2$. We only need the update-case $y_i \left(\boldsymbol{\mu} \cdot \boldsymbol{x}_i\right) - \epsilon > 0$. Apply the update $\boldsymbol{\mu} = \boldsymbol{\mu}_i + \frac{y_i\left(\boldsymbol{\mu}_i \cdot \boldsymbol{x}_i\right) - \epsilon}{0.5\alpha^{-1} - \boldsymbol{x}_i^{\top} \Sigma_i \left(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i\right)} \cdot y_i \Sigma_i \left(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i\right)$ and $\sqrt{u_i} \equiv \left\|\Upsilon \boldsymbol{x}_i\right\|$.

$$\phi\sqrt{u_i} = \left\lfloor y_i \left(\boldsymbol{\mu} \cdot \boldsymbol{x}_i\right) - \epsilon \right\rceil^2 = \left(y_i \left(\boldsymbol{\mu} \cdot \boldsymbol{x}_i\right) - \epsilon\right)^2$$

$$\phi\sqrt{u_i} = \left(y_i \boldsymbol{\mu}_i \cdot \boldsymbol{x}_i - \epsilon + \frac{y_i \left(\boldsymbol{\mu}_i \cdot \boldsymbol{x}_i\right) - \epsilon}{0.5\alpha^{-1} - \boldsymbol{x}_i^{\top} \Sigma_i \left(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i\right)} \cdot \boldsymbol{x}_i^{\top} \Sigma_i \left(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i\right)\right)^2$$

Suppose $g(\alpha) = \left(A + \frac{AC}{0.5\alpha^{-1} - C}\right)^2$, with $(A, C, \alpha_0) = \left(y_i \left(\boldsymbol{\mu}_i \cdot \boldsymbol{x}_i\right) - \epsilon, \boldsymbol{x}_i^{\top} \Sigma_i \left(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i\right), 0\right)$ and use Taylor expansion $g(\alpha) \approx g(\alpha_0) + g^{'}(\alpha_0)(\alpha - \alpha_0)$. It follows that $\left(g(0), g^{'}(0)\right) = \left(A^2, 4A^2C\right)$, thus $\phi\sqrt{u_i} = g(\alpha) \approx A^2 + 4A^2C\alpha \Rightarrow \left(\lambda, \lambda^{'}\right) \approx \left(\left(y_i \left(\boldsymbol{\mu}_i \cdot \boldsymbol{x}_i\right) - \epsilon\right)^2, 4\lambda \boldsymbol{x}_i^{\top} \Sigma_i \left(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i\right)\right)$. □

Lemma 26. $\Sigma_{i+1}^{-1} \blacktriangleright \hbar_{2[\ln]}$

$\equiv$ Lemma 6. □

Lemma 27. $\Sigma_{i+1} \blacktriangleright \hbar_{2[\ln]}$

$\equiv$ Lemma 7. □

Lemma 28. $\sqrt{u_i} \blacktriangleright \hbar_{2[\ln]}$

$\equiv$ Lemma 8. □

Lemma 29. $\boldsymbol{\mu}_{i+1} \blacktriangleright \hbar_{2[\ln]}$

Similar to Lemma 24, $\boldsymbol{\mu} = \boldsymbol{\mu}_i + \alpha \hbar_2^{'} f^{'} y_i \Upsilon_i^2 \left(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i\right)$ with two cases, $y_i \ln\left(\boldsymbol{\mu} \cdot \boldsymbol{x}_i\right) - \epsilon\, [>]\, [\leq]\, 0$.

Case $[>]$: $\hbar_2^{'}(.) = 2\left(y_i \ln\left(\boldsymbol{\mu} \cdot \boldsymbol{x}_i\right) - \epsilon\right)$; use $\ln\left(\boldsymbol{\mu} \cdot \boldsymbol{x}_i\right) \approx \ln\left(\boldsymbol{\mu}_i \cdot \boldsymbol{x}_i\right) + \frac{\left(\boldsymbol{\mu} - \boldsymbol{\mu}_i\right) \cdot \boldsymbol{x}_i}{\boldsymbol{\mu}_i \cdot \boldsymbol{x}_i} \Rightarrow f^{'}(.) = \frac{1}{\boldsymbol{\mu}_i \cdot \boldsymbol{x}_i}$ and $\Upsilon_i^2 = \Sigma_i$,

$$\boldsymbol{\mu} \approx \boldsymbol{\mu}_i + 2\alpha \left(y_i \left(\ln\left(\boldsymbol{\mu}_i \cdot \boldsymbol{x}_i\right) + \frac{\left(\boldsymbol{\mu} - \boldsymbol{\mu}_i\right) \cdot \boldsymbol{x}_i}{\boldsymbol{\mu}_i \cdot \boldsymbol{x}_i}\right) - \epsilon\right) \frac{y_i}{\boldsymbol{\mu}_i \cdot \boldsymbol{x}_i} \Sigma_i \left(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i\right)$$

$$\frac{\left(\boldsymbol{\mu} - \boldsymbol{\mu}_i\right) y_i \boldsymbol{x}_i}{\boldsymbol{\mu}_i \cdot \boldsymbol{x}_i} \approx 2\alpha \frac{\boldsymbol{x}_i^{\top} \Sigma_i \left(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i\right)}{\left(\boldsymbol{\mu}_i \cdot \boldsymbol{x}_i\right)^2} \left(y_i \ln\left(\boldsymbol{\mu}_i \cdot \boldsymbol{x}_i\right) - \epsilon + \frac{\left(\boldsymbol{\mu} - \boldsymbol{\mu}_i\right) y_i \boldsymbol{x}_i}{\boldsymbol{\mu}_i \cdot \boldsymbol{x}_i}\right)$$

Write $X = \frac{\left(\boldsymbol{\mu} - \boldsymbol{\mu}_i\right) y_i \boldsymbol{x}_i}{\boldsymbol{\mu}_i \cdot \boldsymbol{x}_i}, C = y_i \ln\left(\boldsymbol{\mu}_i \cdot \boldsymbol{x}_i\right) - \epsilon, S = 2\alpha \frac{\boldsymbol{x}_i^{\top} \Sigma_i \left(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i\right)}{\left(\boldsymbol{\mu}_i \cdot \boldsymbol{x}_i\right)^2}$,
hence $X = S(C + X) = \frac{SC}{1-S}$ and

$$\left(\boldsymbol{\mu}, C + X\right) \approx \left(\boldsymbol{\mu}_i + 2\alpha (C + X) \cdot \frac{y_i}{\boldsymbol{\mu}_i \cdot \boldsymbol{x}_i} \Sigma_i \left(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i\right), \frac{C}{1-S}\right)$$

Case $[\leq]$: $\hbar_2^{'}(.) = 0 \Rightarrow \left(\boldsymbol{\mu}, C + X\right) = \left(\boldsymbol{\mu}_i + 2\alpha (C + X) \cdot \frac{y_i}{\boldsymbol{\mu}_i \cdot \boldsymbol{x}_i} \Sigma_i \left(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i\right), 0\right)$.

We can conclude with the update $\boldsymbol{\mu}_{i+1} = \boldsymbol{\mu} \approx \boldsymbol{\mu}_i + 2\alpha \lfloor C + X \rfloor y_i \Sigma_i (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i)$

$$\boldsymbol{\mu}_{i+1} \approx \boldsymbol{\mu}_i + \left\lfloor \frac{y_i \ln(\boldsymbol{\mu}_i \cdot \boldsymbol{x}_i) - \epsilon}{0.5\alpha^{-1} - \frac{\boldsymbol{x}_i^\top \Sigma_i (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i)}{(\boldsymbol{\mu}_i \cdot \boldsymbol{x}_i)^2}} \right\rfloor \frac{y_i}{\boldsymbol{\mu}_i \cdot \boldsymbol{x}_i} \Sigma_i (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i)$$

□

Lemma 30. $\alpha \blacktriangleright \hbar_{2[\ln]}$

Similar to Lemma 25, $\alpha = \left\lfloor \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \right\rfloor$ where $(a, b, c) = \left(\lambda' \left(\lambda' + \upsilon_i \phi^2\right), 2\lambda \left(\lambda' + \frac{\upsilon_i \phi^2}{2}\right), \lambda^2 - \upsilon_i \phi^2\right)$.

To find $\left(\lambda, \lambda'\right)$, use binding constraint $0 \leq \phi \|\Upsilon \boldsymbol{x}_i\| = \hbar_{2[ln]} \Rightarrow \phi \|\Upsilon \boldsymbol{x}_i\| = \lfloor y_i \ln(\boldsymbol{\mu} \cdot \boldsymbol{x}_i) - \epsilon \rfloor^2$. We only need the update-case $y_i \ln(\boldsymbol{\mu} \cdot \boldsymbol{x}_i) - \epsilon > 0$. Apply the update

$$\boldsymbol{\mu} = \boldsymbol{\mu}_i + \frac{y_i \ln(\boldsymbol{\mu}_i \cdot \boldsymbol{x}_i) - \epsilon}{0.5\alpha^{-1} - \frac{\boldsymbol{x}_i^\top \Sigma_i (\boldsymbol{x}_i - \overline{\boldsymbol{x}})}{(\boldsymbol{\mu}_i \cdot \boldsymbol{x}_i)^2}} \cdot \frac{y_i}{\boldsymbol{\mu}_i \cdot \boldsymbol{x}_i} \Sigma_i (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i)$$

and $\sqrt{u_i} \equiv \|\Upsilon \boldsymbol{x}_i\|$ to have $\phi\sqrt{u_i} = \lfloor y_i \ln(\boldsymbol{\mu} \cdot \boldsymbol{x}_i) - \epsilon \rfloor^2 = (y_i \ln(\boldsymbol{\mu} \cdot \boldsymbol{x}_i) - \epsilon)^2$.

Use the approximation $y_i \ln(\boldsymbol{\mu} \cdot \boldsymbol{x}_i) - \epsilon \approx y_i \left(\ln(\boldsymbol{\mu}_i \cdot \boldsymbol{x}_i) + \frac{(\boldsymbol{\mu} - \boldsymbol{\mu}_i) \cdot \boldsymbol{x}_i}{\boldsymbol{\mu}_i \cdot \boldsymbol{x}_i}\right) - \epsilon$,

$$\begin{aligned} \phi\sqrt{u_i} &\approx \left(y_i \left(\ln(\boldsymbol{\mu}_i \cdot \boldsymbol{x}_i) + \frac{(\boldsymbol{\mu} - \boldsymbol{\mu}_i) \cdot \boldsymbol{x}_i}{\boldsymbol{\mu}_i \cdot \boldsymbol{x}_i}\right) - \epsilon\right)^2 \\ &\approx \left(y_i \ln(\boldsymbol{\mu}_i \cdot \boldsymbol{x}_i) - \epsilon + \frac{y_i \ln(\boldsymbol{\mu}_i \cdot \boldsymbol{x}_i) - \epsilon}{0.5\alpha^{-1} - \frac{\boldsymbol{x}_i^\top \Sigma_i (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i)}{(\boldsymbol{\mu}_i \cdot \boldsymbol{x}_i)^2}} \cdot \frac{\boldsymbol{x}_i^\top \Sigma_i (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i)}{(\boldsymbol{\mu}_i \cdot \boldsymbol{x}_i)^2}\right)^2 \end{aligned}$$

Similar to Lemma 25, with $(A, C, \alpha_0) = \left(y_i \ln(\boldsymbol{\mu}_i \cdot \boldsymbol{x}_i) - \epsilon, \frac{\boldsymbol{x}_i^\top \Sigma_i (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i)}{(\boldsymbol{\mu}_i \cdot \boldsymbol{x}_i)^2}, 0\right)$; one can show $\phi\sqrt{u_i} \approx A^2 + 4A^2 C\alpha \Rightarrow \left(\lambda, \lambda'\right) \approx \left((y_i \ln(\boldsymbol{\mu}_i \cdot \boldsymbol{x}_i) - \epsilon)^2, 4\lambda \frac{\boldsymbol{x}_i^\top \Sigma_i (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i)}{(\boldsymbol{\mu}_i \cdot \boldsymbol{x}_i)^2}\right)$. □

References

[Beh00] E. Behrends. *Introduction to Markov Chains with Special Emphasis on Rapid Mixing.* Advanced Lecture Notes in Mathematics, Vieweg Verlag, 2000.

[BCG05] N. Cesa-Bianchi, A. Conconi, and C. Gentile. A second-order perceptron algorithm. *SIAM Journal of Commutation*, 34(3): 640–668, 2005.

[CY11] Y. Chen and X. Ye. Projection onto a simplex. Department of Mathematics, University of Florida, 2011.

[CDR07] J.F. Coeurjolly, R. Drouilhet, and J.F. Robineau. Normalized information-based divergences. *Problems of Information Transmission*, 43(3): 167–189, 2007.

[CDF08] K. Crammer, M. Dredze, and F. Pereira. Exact convex confidence-weighted learning. *Neural Information Processing Systems (NIPS)*, 2008.

[Dav06] R. Davidson. Stochastic dominance. Department of Economics, McGill University, 2006.

[LCLMV04] M. Li, X. Chen, X. Li, B. Ma and P.M.B. Vitányi. The similarity metric. *IEEE Trans. Inform. Theory*, 50(12): 3250–3264, 2004.

[LHZG11] B. Li, S.C.H. Hoi, P. Zhao, and V. Gopalkrishnan. Confidence weighted mean reversion strategy for on-line portfolio selection. School of Computer Engineering, Nanyang Technological University, 2011.

[OC09] F. Orabona and and K. Crammer. New adaptive algorithms for online classification. *Neural Information Processing Systems (NIPS)*, 2010.

[PP08] K.B. Petersen and M.S. Pedersen. *The Matrix Cookbook*, 2008.

[Sni10] M. Sniedovich. *Dynamic Programming: Foundations and Principles, 2nd ed.,* CRC Press, 2010.