

MPRA

Munich Personal RePEc Archive

Kullback-Leibler Simplex

Kangpenkae, Popon

June 2012

Online at <https://mpra.ub.uni-muenchen.de/39494/>
MPRA Paper No. 39494, posted 17 Jun 2012 00:49 UTC

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

Inversion [PP08] [146].

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

for our application,

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

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

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\mu}} (\boldsymbol{\mu}_i - \boldsymbol{\mu}) \Upsilon_i^{-2} (\boldsymbol{\mu}_i - \boldsymbol{\mu}) &= 2\Upsilon_i^{-2} (\boldsymbol{\mu} - \boldsymbol{\mu}_i) \\ \frac{\partial}{\partial \Upsilon} \ln(\det \Upsilon^2) &= 2\Upsilon^{-1} \\ \frac{\partial}{\partial \Upsilon} \text{Tr}(\Upsilon_i^{-2} \Upsilon^2) &= \Upsilon_i^{-2} \Upsilon + \Upsilon \Upsilon_i^{-2} \\ \frac{\partial}{\partial \Upsilon} \mathbf{x}_i^\top \Upsilon^2 \mathbf{x}_i &= \mathbf{x}_i \mathbf{x}_i^\top \Upsilon + \Upsilon \mathbf{x}_i \mathbf{x}_i^\top = \frac{\partial}{\partial \Upsilon} \|\Upsilon \mathbf{x}_i\|^2 \\ \frac{\partial}{\partial \Upsilon} \|\Upsilon \mathbf{x}_i\| &= \frac{\mathbf{x}_i \mathbf{x}_i^\top \Upsilon + \Upsilon \mathbf{x}_i \mathbf{x}_i^\top}{2\|\Upsilon \mathbf{x}_i\|} = \frac{\mathbf{x}_i \mathbf{x}_i^\top \Upsilon + \Upsilon \mathbf{x}_i \mathbf{x}_i^\top}{2\sqrt{\mathbf{x}_i^\top \Upsilon^2 \mathbf{x}_i}} \end{aligned}$$

KL divergence.

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

3. APPROXIMATION

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

As KL simplex solution in Δ does not have a closed form, the approximation will start with $\overleftrightarrow{\Delta}$,

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

subject to $\hbar(y_i f(\boldsymbol{\mu} \cdot \mathbf{x}_i) - \epsilon) \geq \phi \sqrt{\mathbf{x}_i^\top \Sigma \mathbf{x}_i}$, $y_i \in \{-1, 1\}$, and $\boldsymbol{\mu} \in \overleftrightarrow{\Delta}$.

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

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

We consider [normal, hinge, hinge²] constraint (see section Section 5), with two flavors:

{linear, logarithm} = {[ln], [ln]} $\ni f(\cdot)$. Let $\Sigma_{[i]} = \Upsilon_{[i]}^2$ where $\Upsilon_{[i]}$ has SSD, the \hbar -Lagrangian is

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

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

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

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

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

where $\beta = \frac{\alpha\phi}{\sqrt{u_i} + \alpha\phi v_i}$, $(u_i, v_i) \equiv (\mathbf{x}_i^\top \Sigma_{i+1} \mathbf{x}_i, \mathbf{x}_i^\top \Sigma_i \mathbf{x}_i)$.Lemma 3. $\sqrt{u_i} \blacktriangleright \hbar_{\emptyset[ln]}$

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

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

$$\boldsymbol{\mu}_{i+1} = \boldsymbol{\mu}_i + \alpha y_i \Sigma_i (\mathbf{x}_i - \bar{\mathbf{x}}_i)$$

where $\bar{\mathbf{x}} = \bar{\mathbf{x}} \mathbf{1} \equiv \frac{\mathbf{1}^\top \Sigma_i \mathbf{x}_i}{\mathbf{1}^\top \Sigma_i \mathbf{1}} \mathbf{1}$ Lemma 5. $\alpha \blacktriangleright \hbar_{\emptyset[ln]}$, $\alpha = \left\lfloor \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \right\rfloor$ such that

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

$$(\lambda, \lambda') = \left(y_i (\boldsymbol{\mu}_i \cdot \mathbf{x}_i) - \epsilon, \mathbf{x}_i^\top \Sigma_i (\mathbf{x}_i - \bar{\mathbf{x}}_i) \right)$$

3.2.2. Logarithm : $\hbar_{\emptyset[ln]}$.Lemma 6. $\Sigma_{i+1}^{-1} \blacktriangleright \hbar_{\emptyset[ln]} \equiv$ Lemma 1.Lemma 7. $\Sigma_{i+1} \blacktriangleright \hbar_{\emptyset[ln]} \equiv$ Lemma 2.Lemma 8. $\sqrt{u_i} \blacktriangleright \hbar_{\emptyset[ln]} \equiv$ Lemma 3.Lemma 9. $\boldsymbol{\mu}_{i+1} \blacktriangleright \hbar_{\emptyset[ln]}$

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

where $\bar{\mathbf{x}} = \bar{\mathbf{x}} \mathbf{1} \equiv \frac{\mathbf{1}^\top \Sigma_i \mathbf{x}_i}{\mathbf{1}^\top \Sigma_i \mathbf{1}} \mathbf{1}$.Lemma 10. $\alpha \blacktriangleright \hbar_{\emptyset[ln]}$, $\alpha = \left\lfloor \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \right\rfloor$ such that

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

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

3.3. **[hinge]**, \hat{h}_1 and **[hinge²]**, \hat{h}_2 .3.3.1. *Linear* : $\hat{h}_{1[l_n]}$, $\hat{h}_{2[l_n]}$.

$$\Sigma_{i+1}^{-1} \blacktriangleright \hat{h}_{[1,2][l_n]}, \text{ Lemma 11} \equiv \text{Lemma 21} \equiv \text{Lemma 1}, \Sigma_{i+1}^{-1} \blacktriangleright \hat{h}_{\emptyset[l_n]}$$

$$\Sigma_{i+1} \blacktriangleright \hat{h}_{[1,2][l_n]}, \text{ Lemma 12} \equiv \text{Lemma 22} \equiv \text{Lemma 2}, \Sigma_{i+1} \blacktriangleright \hat{h}_{\emptyset[l_n]}$$

$$\sqrt{u_i} \blacktriangleright \hat{h}_{[1,2][l_n]}, \text{ Lemma 13} \equiv \text{Lemma 23} \equiv \text{Lemma 3}, \sqrt{u_i} \blacktriangleright \hat{h}_{\emptyset[l_n]}$$

$$\text{Lemma 14. } \boldsymbol{\mu}_{i+1} \blacktriangleright \hat{h}_{1[l_n]}$$

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

$$\text{where } \bar{\mathbf{x}}_i = \bar{\mathbf{x}}_i \mathbf{1} \equiv \frac{\mathbf{1}^\top \Sigma_i \mathbf{x}_i}{\mathbf{1}^\top \Sigma_i \mathbf{1}} \mathbf{1}$$

$$\text{Lemma 15. } \alpha \blacktriangleright \hat{h}_{1[l_n]}, \alpha = \left\lfloor \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \right\rfloor \text{ such that}$$

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

$$(\lambda, \lambda') = (y_i (\boldsymbol{\mu}_i \cdot \mathbf{x}_i) - \epsilon, \mathbf{x}_i^\top \Sigma_i (\mathbf{x}_i - \bar{\mathbf{x}}_i))$$

$$\text{Lemma 24. } \boldsymbol{\mu}_{i+1} \blacktriangleright \hat{h}_{2[l_n]}$$

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

$$\text{where } \bar{\mathbf{x}}_i = \bar{\mathbf{x}}_i \mathbf{1} \equiv \frac{\mathbf{1}^\top \Sigma_i \mathbf{x}_i}{\mathbf{1}^\top \Sigma_i \mathbf{1}} \mathbf{1}$$

$$\text{Lemma 25. } \alpha \blacktriangleright \hat{h}_{2[l_n]}, \alpha = \left\lfloor \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \right\rfloor \text{ such that}$$

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

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

3.3.2. *Logarithm* : $\hat{h}_{1[l_n]}$, $\hat{h}_{2[l_n]}$.

$$\Sigma_{i+1}^{-1} \blacktriangleright \hat{h}_{[1,2][l_n]}, \text{ Lemma 16} \equiv \text{Lemma 26} \equiv \text{Lemma 6}, \Sigma_{i+1}^{-1} \blacktriangleright \hat{h}_{\emptyset[l_n]}$$

$$\Sigma_{i+1} \blacktriangleright \hat{h}_{[1,2][l_n]}, \text{ Lemma 17} \equiv \text{Lemma 27} \equiv \text{Lemma 7}, \Sigma_{i+1} \blacktriangleright \hat{h}_{\emptyset[l_n]}$$

$$\sqrt{u_i} \blacktriangleright \hat{h}_{[1,2][l_n]}, \text{ Lemma 18} \equiv \text{Lemma 28} \equiv \text{Lemma 8}, \sqrt{u_i} \blacktriangleright \hat{h}_{\emptyset[l_n]}$$

$$\text{Lemma 19. } \boldsymbol{\mu}_{i+1} \blacktriangleright \hat{h}_{1[l_n]}$$

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

$$\text{where } \bar{\mathbf{x}}_i = \bar{\mathbf{x}}_i \mathbf{1} \equiv \frac{\mathbf{1}^\top \Sigma_i \mathbf{x}_i}{\mathbf{1}^\top \Sigma_i \mathbf{1}} \mathbf{1}$$

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

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

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

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

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

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

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

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

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

4. IMPLEMENTATION

4.1. Results in section 3. is valid for the $\overset{\leftrightarrow}{\Delta}$ simplex. A more common constraint is Δ simplex; however the close-form solution is not possible with this simplex. Projecting simplex $\overset{\leftrightarrow}{\Delta}$ on Δ 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 $\Delta_i \in \{1 \dots m\}$ be a set of *nSD-effective* simplex, each $\Delta_i \in [0, 1]^N$. Connect m subsimplex into a vector in $[0, 1]^{m \cdot N}$; apply simplex projection to the vector. The result is simplex $\Delta \in [0, 1]^{m \cdot N}$; overlay simplex Δ , i.e. slot Δ 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.

\ddagger *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 $\frac{F(\mathbf{w} \cdot \mathbf{x}_i) - \mu_{F(\mathbf{w} \cdot \mathbf{x}_i)}}{\sigma_{F(\mathbf{w} \cdot \mathbf{x}_i)}} = Z_{\Phi - \text{cdf}}$; consider a generic confidence constraint $\Pr(F(\mathbf{w} \cdot \mathbf{x}_i) \geq 0) \geq \eta \equiv \Phi(\phi)$.

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

$$\frac{-\mu_{F(\mathbf{w} \cdot \mathbf{x}_i)}}{\sigma_{F(\mathbf{w} \cdot \mathbf{x}_i)}} \leq \Phi^{-1}(1 - \eta) = -\Phi^{-1}(\eta) \Rightarrow \mu_{F(\mathbf{w} \cdot \mathbf{x}_i)} \geq \Phi^{-1}(\eta) \sigma_{F(\mathbf{w} \cdot \mathbf{x}_i)} = \phi \sigma_{F(\mathbf{w} \cdot \mathbf{x}_i)}$$

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

5.2. The approximating property of [normal, hinge, hinge²] confidence. Define [normal, hinge, hinge²] function as follows,

$$\begin{aligned} \text{normal: } \hat{h}_{\emptyset[f]} &\in \{\hat{h}_{\emptyset[l_n]}, \hat{h}_{\emptyset[l_n]}\} \equiv \{y_i(\boldsymbol{\mu} \cdot \mathbf{x}_i) - \epsilon, y_i \ln(\boldsymbol{\mu} \cdot \mathbf{x}_i) - \epsilon\} \\ \text{hinge: } \hat{h}_{1[f]} &\in \{\hat{h}_{1[l_n]}, \hat{h}_{1[l_n]}\} \equiv \{\lfloor y_i(\boldsymbol{\mu} \cdot \mathbf{x}_i) - \epsilon \rfloor, \lfloor y_i \ln(\boldsymbol{\mu} \cdot \mathbf{x}_i) - \epsilon \rfloor\} \\ \text{hinge}^2: \hat{h}_{2[f]} &\in \{\hat{h}_{2[l_n]}, \hat{h}_{2[l_n]}\} \equiv \{\lfloor y_i(\boldsymbol{\mu} \cdot \mathbf{x}_i) - \epsilon \rfloor^2, \lfloor y_i \ln(\boldsymbol{\mu} \cdot \mathbf{x}_i) - \epsilon \rfloor^2\} \end{aligned}$$

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

normal: $\hat{h}_{\emptyset[l_n]}$ is exact; $\hat{h}_{\emptyset[l_n]}$ is approximate

$$\begin{aligned} F(\mathbf{w} \cdot \mathbf{x}_i) &= y_i(\mathbf{w} \cdot \mathbf{x}_i) - \epsilon \Rightarrow (\mu_{F(\mathbf{w} \cdot \mathbf{x}_i)}, \sigma_{F(\mathbf{w} \cdot \mathbf{x}_i)}^2) = (y_i(\boldsymbol{\mu} \cdot \mathbf{x}_i) - \epsilon, \sigma_{\mathbf{w} \cdot \mathbf{x}_i} = \mathbf{x}_i^\top \Sigma \mathbf{x}_i) \\ F(\mathbf{w} \cdot \mathbf{x}_i) &= y_i \ln(\mathbf{w} \cdot \mathbf{x}_i) - \epsilon \Rightarrow \mu_{F(\mathbf{w} \cdot \mathbf{x}_i)} \approx y_i \ln(\boldsymbol{\mu} \cdot \mathbf{x}_i) \end{aligned}$$

hinge: $\hat{h}_{1[l_n][l_n]}$ is approximate

$$\begin{aligned} F(\mathbf{w} \cdot \mathbf{x}_i) &= \lfloor y_i(\mathbf{w} \cdot \mathbf{x}_i) - \epsilon \rfloor \Rightarrow \mu_{F(\mathbf{w} \cdot \mathbf{x}_i)} \approx \lfloor y_i(\boldsymbol{\mu} \cdot \mathbf{x}_i) - \epsilon \rfloor \\ F(\mathbf{w} \cdot \mathbf{x}_i) &= \lfloor y_i \ln(\mathbf{w} \cdot \mathbf{x}_i) - \epsilon \rfloor \Rightarrow \mu_{F(\mathbf{w} \cdot \mathbf{x}_i)} \approx \lfloor y_i \ln(\boldsymbol{\mu} \cdot \mathbf{x}_i) - \epsilon \rfloor \end{aligned}$$

hinge²: $\hat{h}_{2[l_n][l_n]}$ is approximate

$$\begin{aligned} F(\mathbf{w} \cdot \mathbf{x}_i) &= \lfloor y_i(\mathbf{w} \cdot \mathbf{x}_i) - \epsilon \rfloor^2 \Rightarrow \mu_{F(\mathbf{w} \cdot \mathbf{x}_i)} \approx \lfloor y_i(\boldsymbol{\mu} \cdot \mathbf{x}_i) - \epsilon \rfloor^2 \\ F(\mathbf{w} \cdot \mathbf{x}_i) &= \lfloor y_i \ln(\mathbf{w} \cdot \mathbf{x}_i) - \epsilon \rfloor^2 \Rightarrow \mu_{F(\mathbf{w} \cdot \mathbf{x}_i)} \approx \lfloor y_i \ln(\boldsymbol{\mu} \cdot \mathbf{x}_i) - \epsilon \rfloor^2 \end{aligned}$$

APPENDIX

Lemma 1. $\Sigma_{i+1}^{-1} \blacktriangleright \hat{h}_{\emptyset[l_n]}$

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

Υ^{-1} update condition is,

$$\Upsilon^{-1} = \frac{1}{2} \Upsilon_i^{-2} \Upsilon + \frac{1}{2} \Upsilon \Upsilon_i^{-2} + \alpha \phi \frac{\mathbf{x}_i \mathbf{x}_i^\top \Upsilon}{2\sqrt{\mathbf{x}_i^\top \Upsilon^2 \mathbf{x}_i}} + \alpha \phi \frac{\Upsilon \mathbf{x}_i \mathbf{x}_i^\top}{2\sqrt{\mathbf{x}_i^\top \Upsilon^2 \mathbf{x}_i}} \quad [\Upsilon^{-1}]$$

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

$$\Upsilon^{-2} \equiv \Upsilon_{i+1}^{-2} = \Upsilon_i^{-2} + \alpha \phi \frac{\mathbf{x}_i \mathbf{x}_i^\top}{\sqrt{\mathbf{x}_i^\top \Upsilon^2 \mathbf{x}_i}} \quad [\Upsilon^{-2}]$$

which yields

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

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

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

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

□

Lemma 2. $\Sigma_{i+1} \blacktriangleright \hat{h}_{\emptyset[l_n]}$

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

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

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

□

Lemma 3. $\sqrt{u_i} \blacktriangleright \hat{h}_{\emptyset[l_n]}$

$$\Sigma_{i+1} = \Sigma_i - \frac{\alpha\phi\Sigma_i\mathbf{x}_i\mathbf{x}_i^\top\Sigma_i}{\sqrt{u_i} + \alpha\phi v_i} \Rightarrow \mathbf{x}_i^\top\Sigma_{i+1}\mathbf{x}_i = \mathbf{x}_i^\top\Sigma_i\mathbf{x}_i - \frac{\alpha\phi(\mathbf{x}_i^\top\Sigma_i\mathbf{x}_i)(\mathbf{x}_i^\top\Sigma_i\mathbf{x}_i)}{\sqrt{u_i} + \alpha\phi v_i}$$

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

□

Lemma 4. $\boldsymbol{\mu}_{i+1} \blacktriangleright \hat{h}_{\emptyset[l_n]}$

$$\frac{\partial}{\partial \boldsymbol{\mu}} \mathcal{L} = 0 = \Upsilon_i^{-2}(\boldsymbol{\mu} - \boldsymbol{\mu}_i) - \alpha\hat{h}'_{\emptyset} f' y_i \mathbf{x}_i + \rho \mathbf{1}; \quad \frac{\partial}{\partial \rho} \mathcal{L} = 0 = \boldsymbol{\mu} \cdot \mathbf{1} - 1$$

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

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

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

use $\hat{h}'_{\emptyset}(\cdot) = 1$, $f'(\cdot) = 1$ and $\Upsilon_i^2 = \Sigma_i$ to have $\boldsymbol{\mu}_{i+1} = \boldsymbol{\mu} = \boldsymbol{\mu}_i + \alpha y_i \Sigma_i (\mathbf{x}_i - \bar{\mathbf{x}}_i)$

□

Lemma 5. $\alpha \blacktriangleright \bar{h}_{\emptyset[l_n]}$

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

To find (λ, λ') , use binding constraint $\phi \|\Upsilon \mathbf{x}_i\| = \bar{h}_{\emptyset[l_n]} \Rightarrow \phi \|\Upsilon \mathbf{x}_i\| = y_i (\boldsymbol{\mu} \cdot \mathbf{x}_i) - \epsilon$. Apply the update $\boldsymbol{\mu} = \boldsymbol{\mu}_i + \alpha y_i \Sigma_i (\mathbf{x}_i - \bar{\mathbf{x}}_i)$ and $\sqrt{u_i} \equiv \|\Upsilon \mathbf{x}_i\|$,

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

$$\text{i.e. } (\lambda, \lambda') = (y_i \boldsymbol{\mu}_i \cdot \mathbf{x}_i - \epsilon, \mathbf{x}_i^\top \Sigma_i (\mathbf{x}_i - \bar{\mathbf{x}}_i)). \quad \square$$

Lemma 6. $\Sigma_{i+1}^{-1} \blacktriangleright \bar{h}_{\emptyset[l_n]}$

\equiv Lemma 1. □

Lemma 7. $\Sigma_{i+1} \blacktriangleright \bar{h}_{\emptyset[l_n]}$

\equiv Lemma 2. □

Lemma 8. $\sqrt{u_i} \blacktriangleright \bar{h}_{\emptyset[l_n]}$

\equiv Lemma 3. □

Lemma 9. $\boldsymbol{\mu}_{i+1} \blacktriangleright \bar{h}_{\emptyset[l_n]}$

Similar to Lemma 4, $\boldsymbol{\mu} = \boldsymbol{\mu}_i + \alpha \bar{h}'_0 f' y_i \Upsilon_i^2 (\mathbf{x}_i - \bar{\mathbf{x}}_i)$; use $\bar{h}'_0(\cdot) = 1$; $\ln(\boldsymbol{\mu} \cdot \mathbf{x}_i) \approx \ln(\boldsymbol{\mu}_i \cdot \mathbf{x}_i) + \frac{(\boldsymbol{\mu} - \boldsymbol{\mu}_i) \cdot \mathbf{x}_i}{\boldsymbol{\mu}_i \cdot \mathbf{x}_i} \Rightarrow f'(\cdot) = \frac{1}{\boldsymbol{\mu}_i \cdot \mathbf{x}_i}$ and $\Upsilon_i^2 = \Sigma_i$, which gives $\boldsymbol{\mu}_{i+1} = \boldsymbol{\mu} \approx \boldsymbol{\mu}_i + \frac{\alpha y_i}{\boldsymbol{\mu}_i \cdot \mathbf{x}_i} \Sigma_i (\mathbf{x}_i - \bar{\mathbf{x}}_i)$ □

Lemma 10. $\alpha \blacktriangleright \bar{h}_{\emptyset[l_n]}$

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

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

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

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

Lemma 11. $\Sigma_{i+1}^{-1} \blacktriangleright \bar{h}_{1[l_n]}$

\equiv Lemma 1. □

Lemma 12. $\Sigma_{i+1} \blacktriangleright \bar{h}_{1[l_n]}$

\equiv Lemma 2. □

Lemma 13. $\sqrt{u_i} \blacktriangleright \hat{h}_{1[l_n]}$

\equiv Lemma 3. □

Lemma 14. $\boldsymbol{\mu}_{i+1} \blacktriangleright \hat{h}_{1[l_n]}$

Similar to Lemma Lemma 4, $\boldsymbol{\mu} = \boldsymbol{\mu}_i + \alpha \hat{h}'_1 f' y_i \Upsilon_i^2 (\mathbf{x}_i - \bar{\mathbf{x}}_i)$. There are two cases, $y_i (\boldsymbol{\mu} \cdot \mathbf{x}_i) - \epsilon [>] [\leq] 0$.

Case [$>$]: $\hat{h}'_1 (\cdot) = 1, f' (\cdot) = 1$ and $\Upsilon_i^2 = \Sigma_i, \Rightarrow \boldsymbol{\mu}_{i+1} = \boldsymbol{\mu} = \boldsymbol{\mu}_i + \alpha y_i \Sigma_i (\mathbf{x}_i - \bar{\mathbf{x}}_i)$

Case [\leq]: $\hat{h}'_1 (\cdot) = 0 \Rightarrow \boldsymbol{\mu}_{i+1} = \boldsymbol{\mu} = \boldsymbol{\mu}_i$

With some manipulation we find a $\boldsymbol{\mu}$ -update

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

□

Lemma 15. $\alpha \blacktriangleright \hat{h}_{1[l_n]}$

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

To find (λ, λ') , use binding constraint $\phi \|\Upsilon \mathbf{x}_i\| = \hat{h}_{1[l_n]} \Rightarrow \phi \|\Upsilon \mathbf{x}_i\| = \lfloor y_i (\boldsymbol{\mu} \cdot \mathbf{x}_i) - \epsilon \rfloor$. We only need the update-case $y_i (\boldsymbol{\mu} \cdot \mathbf{x}_i) - \epsilon > 0$. Apply the update $\boldsymbol{\mu} = \boldsymbol{\mu}_i + \alpha y_i \Sigma_i (\mathbf{x}_i - \bar{\mathbf{x}}_i)$ and $\sqrt{u_i} \equiv \|\Upsilon \mathbf{x}_i\|$,

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

i.e. $(\lambda, \lambda') = (y_i \boldsymbol{\mu}_i \cdot \mathbf{x}_i - \epsilon, \mathbf{x}_i^\top \Sigma_i (\mathbf{x}_i - \bar{\mathbf{x}}_i))$. □

Lemma 16. $\Sigma_{i+1}^{-1} \blacktriangleright \hat{h}_{1[l_n]}$

\equiv Lemma 6. □

Lemma 17. $\Sigma_{i+1} \blacktriangleright \hat{h}_{1[l_n]}$

\equiv Lemma 7. □

Lemma 18. $\sqrt{u_i} \blacktriangleright \hat{h}_{1[l_n]}$

\equiv Lemma 8. □

Lemma 19. $\boldsymbol{\mu}_{i+1} \blacktriangleright \hat{h}_{1[l_n]}$

Similar to Lemma 14, $\boldsymbol{\mu} = \boldsymbol{\mu}_i + \alpha \hat{h}'_1 f' y_i \Upsilon_i^2 (\mathbf{x}_i - \bar{\mathbf{x}}_i)$ with two cases, $y_i \ln (\boldsymbol{\mu} \cdot \mathbf{x}_i) - \epsilon [>] [\leq] 0$.

Case [$>$]: $\hat{h}'_1 (\cdot) = 1, \ln (\boldsymbol{\mu} \cdot \mathbf{x}_i) \approx \ln (\boldsymbol{\mu}_i \cdot \mathbf{x}_i) + \frac{(\boldsymbol{\mu} - \boldsymbol{\mu}_i) \cdot \mathbf{x}_i}{\boldsymbol{\mu}_i \cdot \mathbf{x}_i} \Rightarrow f' (\cdot) = \frac{1}{\boldsymbol{\mu}_i \cdot \mathbf{x}_i}$ and $\Upsilon_i^2 = \Sigma_i$

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

Case [\leq]: $\hat{h}'_1 (\cdot) = 0 \Rightarrow \boldsymbol{\mu}_{i+1} = \boldsymbol{\mu} = \boldsymbol{\mu}_i$

With some manipulation we find a $\boldsymbol{\mu}$ -update

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

□

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

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

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

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

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

\equiv Lemma 1. \square

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

\equiv Lemma 2. \square

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

\equiv Lemma 3. \square

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

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

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

$$\boldsymbol{\mu} = \boldsymbol{\mu}_i + 2\alpha(y_i(\boldsymbol{\mu} \cdot \mathbf{x}_i) - \epsilon) y_i \Sigma_i(\mathbf{x}_i - \bar{\mathbf{x}}_i)$$

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

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

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

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

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

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

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

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

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

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

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

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

\equiv Lemma 6. \square

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

\equiv Lemma 7. \square

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

\equiv Lemma 8. \square

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

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

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

$$\begin{aligned} \boldsymbol{\mu} &\approx \boldsymbol{\mu}_i + 2\alpha \left(y_i \left(\ln(\boldsymbol{\mu}_i \cdot \mathbf{x}_i) + \frac{(\boldsymbol{\mu} - \boldsymbol{\mu}_i) \cdot \mathbf{x}_i}{\boldsymbol{\mu}_i \cdot \mathbf{x}_i} \right) - \epsilon \right) \frac{y_i}{\boldsymbol{\mu}_i \cdot \mathbf{x}_i} \Sigma_i (\mathbf{x}_i - \bar{\mathbf{x}}_i) \\ \frac{(\boldsymbol{\mu} - \boldsymbol{\mu}_i) y_i \mathbf{x}_i}{\boldsymbol{\mu}_i \cdot \mathbf{x}_i} &\approx 2\alpha \frac{\mathbf{x}_i^\top \Sigma_i (\mathbf{x}_i - \bar{\mathbf{x}}_i)}{(\boldsymbol{\mu}_i \cdot \mathbf{x}_i)^2} \left(y_i \ln(\boldsymbol{\mu}_i \cdot \mathbf{x}_i) - \epsilon + \frac{(\boldsymbol{\mu} - \boldsymbol{\mu}_i) y_i \mathbf{x}_i}{\boldsymbol{\mu}_i \cdot \mathbf{x}_i} \right) \end{aligned}$$

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

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

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

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

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

□

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

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

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

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

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

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

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

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

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 Communication*, 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 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.