Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN
ECONOMICS
Ciclo 31

Settore Concorsuale: 13/A1
Settore Scientifico Disciplinare: SECS-P/01

Essays in Minimum Relative Entropy implementations for views processing

Presentata da: Marcello Colasante

Coordinatore Dottorato

Prof.ssa Maria Bigoni

Supervisore

Prof. Massimiliano Marzo

Co-supervisore

Dott. Attilio Meucci

Esame finale anno 2019

Abstract

In this thesis we investigate aspects of the theory of minimum relative entropy models (MRE in the sequel) within the class of exponential-family distributions. We use this technique for an application in portfolio management to compute Bayesian-like statistical features that incorporate fully general views on multivariate markets.

The remainder of this dissertation is structured in three connected chapters.

In Chapter 1 we focus on the parametric implementation of MRE under the normal assumption of the market variables and equality views on linear combinations of their first two moments. Under these special assumptions, we compute the analytical formulations of the Lagrange multipliers that steer the canonical MRE updated distribution within the exponential-family class. This allows to represent the MRE solution through a more parsimonious parametrization and to generalize the current and notable results on the parametric implementation of MRE.

In Chapter 2 we generalize results in Chapter 1 under more flexible (in)equality views on linear combinations of the first two moments of the market variables. To this purpose, we supply the analytical derivatives of the dual Lagrangian objective in order to compute numerically the Lagrange multipliers identifying the MRE solution. Finally, we use this implementation for the construction of quantitative trading strategies based on ranking signals for alpha-generation, the so-called "portfolios from sorts".

In Chapter 3 we address the MRE problem under no assumption on the market variables and (in)equality views on their generalized expectations. In such circumstances the exact computation of the solution is not practically possible. For this reason, we introduce a numerical implementation through iterated Hamiltonian Monte Carlo simulations which efficiently addresses the parameter estimation of the MRE updated distribution within the exponential-family class. This yields a generalization of the non-parametric implementation of MRE that reduces the statistical error of the estimators for a given sample size.

Fully documented code is available on GitHub.

JEL Classification: C1, G11

Keywords: Black-Litterman, Bayesian estimation, Regression, Minimum Relative Entropy, Flexible Probabilities, views, Kullback-Leibler, Hamiltonian Monte Carlo, portfolios from sorts, exponential-family distributions.

General introduction

In this thesis we generalize aspects of the theory of minimum relative entropy in [Meucci, 2008]. We use minimum relative entropy approach to combine traditional econometric techniques for statistical prediction with regularization methods which allow to embed any possible informations in our market models. Such informations may arrive in the form of new data, subjective beliefs, trading signals, extreme scenarios etc. In order to cover all such instances in full generality, we summarize all these informations with the generic nomenclature of "views". See also [Meucci, 2019].

The reason why we need to overlay views is multiple.

As a matter of fact, classical estimates, such as sample means and covariances, are highly unstable, or more precisely *inefficient*, in that they are too sensitive to the observed empirical data [Stein, 1955], [Lehmann and Casella, 1998]. Hence, pure classical estimates are not suitable for portfolio management: they cannot be simply "plugged in" portfolio optimizations, such as optimal portfolio choices a-la "Modern Portfolio Theory" [Markowitz, 1952]. See for instance [DeMiguel et al., 2009] and [Kan and Zhou, 2006].

However, pure Bayesian approaches that only take into account views without considering the observed empirical data can be inaccurate, or more precisely *biased*: the outcomes are concentrated around a value, which may be far from the true property to be estimated.

Processing views in a statistically correct way is not an easy task. Most notable techniques, including regression-based [RiskMetrics, 1996], Bayesian [Aitchison and Dunsmore, 1975] and Black-Litterman approach [Black and Litterman, 1990], can only process *global* informations and often rely on restrictive assumptions (linear approximations, truncations, normality etc.). Hence these approaches are not enough flexible to perform generalized stress-testing and scenario analysis (say on correlations or volatilities).

Here we focus on the minimum relative entropy approach (MRE), also known as principle of minimum discrimination information (MINXENT) or maximum entropy principle (MAXENT) [Jaynes, 1957a], [Jaynes, 1957b], which has been notably applied also in physics, statistics and information theory [Cover and Thomas, 2006], machine learning [Malouf, 2002] and other areas of finance [Hansen and Sargent, 2007], [Glasserman and Xu, 2013], [Avellaneda, 1999], [D'Amico et al., 2003], [Cont, 2007], [Breuer and Csiszar, 2013], [Pezier, 2007], [Friedman et al., 2012].

MRE, generalizes the well-known linear conditioning/regression [Meucci, 2008], the Bayesian methodology [Caticha and Giffin, 2006], the maximum likelihood approach

[Kriz and Talacko, 1968], the Black-Litterman approach [Meucci, 2010], and can be mainly implemented either non-parametrically, or parametrically, as summarized in table below. Refer also to [Meucci, 2008] for more details.

MRE	Parametric	Non-parametric
Model	normal	MC scenarios
Views	linear equalities on exp. and cov.	general (in)equalities on exp.
Solution	analytical	numerical

Table 1: Approaches to MRE

In the parametric approach, the MRE problem is addressed analytically under equality views in the format of linear expectation/covariance-based conditions. However the implementation requires the normality of the market variables.

From the other hand, in the non-parametric approach the MRE problem is addressed numerically via Monte Carlo simulations under (in)equality views in the format of generalized expectation-based conditions. However the statistical precision of the non-parametric implementation becomes low in large dimensions.

The main focus of this thesis is to further extend theoretical and numerical results behind these two implementations.

Outline

This thesis is structured as follows.

Chapter 1

In Chapter 1 we prove the invariance of the MRE solution within the exponential-family class. Then we describe the parametric implementation of the MRE under normality assumption, addressing the problem specifically for the cases of views on linear combinations of i) first moments; ii) second moments; and iii) first two moments of the market variables. Here we also show how it is possible to derive analytically the canonical exponential-family parametrization of the MRE solution when we consider central moment conditions, generalizing the original formulation by [Meucci, 2010].

Chapter 2

In Chapter 2 we generalize the parametric implementation of the MRE introduced in Chapter 1 to the case of more general (in)equality views. In this case, we show how to derive the MRE solution addressing numerically the dual Lagrangian problem. We also supply the analytical derivatives of the objective in order to enhance the precision of such computation. Finally, we use this implementation of MRE to build and backtest a systematic strategy based on ranking views stemming from trading signals.

Chapter 3

In Chapter 3 we describe a numerical implementation of the MRE via iterative Hamiltonian Monte Carlo simulations which allows efficient estimation of the parameters, or Lagrange multipliers, driving the MRE solution under generalized (in)equality views on expectations and arbitrary market model. Then we compare the performance of this approach with the non-parametric implementation of the MRE showing how the iterative implementation can significantly better reduce the estimation error of the Lagrange multipliers.

Background

In this section we briefly introduce the basic notions behind views processing problems, drawing notations and results from [Meucci, 2019]. Refer also to [Cover and Thomas, 2006] for further details.

Base distribution and view variables

In order to start, let us consider the main ingredients:

1. an $\bar{n} \times 1$ random variable $\boldsymbol{X} \equiv (X_1, \ldots, X_{\bar{n}})'$, the **market**, representing the future randomness we want to model, say tomorrow's returns of the stocks in the S&P 500. The market variables are associated with a reference **base distribution**, as represented by their joint probability density function (pdf), which we denote by

$$\boldsymbol{X} \sim \underline{f}_{\boldsymbol{X}}. \tag{1}$$

   The base has to be thought as the initial guess for the true and unknown market distribution and in practice is the outcome of classical estimation techniques (historical, maximum likelihood, GMM etc.). In real applications the number $\bar{n}$ of market variables is *large* (say, of the order of hundreds or thousands).

2. a $\bar{k} \times 1$ random variable $\boldsymbol{Z} \equiv (Z_1, \ldots, Z_{\bar{k}})'$, the factors or **view variables**, on which we have new informations or views that have a potential impact on the market variables, say the return of the S&P 500 index. Without loss of generality, we can assume the view variables to be *endogenous*, in that they can be expressed as a transformation of the market variables

$$\boldsymbol{Z} \equiv \zeta^{view}(\boldsymbol{X}), \tag{2}$$

for a suitable multivariate function $\zeta^{view} : \mathbb{R}^{\bar{n}} \to \mathbb{R}^{\bar{k}}$. Then the base distribution of the market variables (1) naturally implies a (base) distribution for the view variables

$$\boldsymbol{X} \sim \underline{f}_{\boldsymbol{X}} \quad \Rightarrow \quad \boldsymbol{Z} \sim \underline{f}_{\boldsymbol{Z}}, \tag{3}$$

In real applications the view variables are significantly *less* than the market variables (say, of the order of dozens)

$$\bar{k} \ll \bar{n}. \tag{4}$$

Point, distributional and partial views

The views are statements on the view variables $\boldsymbol{Z}$ (2) that potentially perturb the initial state, implied by the base distribution, into a new state, as represented by the conditional distribution given the views, also known as *updated distribution*

$$\begin{array}{ccc} \text{Original state (base)} & \text{Views} & \text{New state (updated)} \\ \underline{f}_{\boldsymbol{Z}}\ (3) & \overset{\text{statements on } \boldsymbol{Z}}{\Rightarrow} & \bar{f}_{\boldsymbol{Z}} \end{array} \tag{5}$$

Such statements are assumed to affect the market variables $\boldsymbol{X}$ in turn

$$\begin{array}{ccc} \text{Original state (base)} & \text{Views} & \text{New state (updated)} \\ \underline{f}_{\boldsymbol{X}}\ (1) & \overset{\text{statements on } \boldsymbol{Z}}{\Rightarrow} & \bar{f}_{\boldsymbol{X}} \end{array} \tag{6}$$

According to the most classical literature in finance and, more in general, in statistics, the simplest views processing problems we usually face are simple statements on the outcomes of the view variables ("How will behave tomorrow's returns of the S&P 500, if the today's index return is 0.01%?").

Such problems can be formalized in the following format.

What is the conditional distribution of the market variables, given a **point view**, i.e. a realization $\boldsymbol{z}^{view}$ of the factors

$$\boldsymbol{X}|\{\boldsymbol{Z} = \boldsymbol{z}^{view}\} \sim ? \tag{7}$$

Views processing problems as in (7) can be easily generalized. To this purpose we first notice that a point view can be interpreted as a statement on the updated distribution $\bar{f}_{\boldsymbol{Z}}$ of the view variables

$$\boldsymbol{Z} = \boldsymbol{z}^{view} \quad \Leftrightarrow \quad \bar{f}_{\boldsymbol{Z}} = \delta^{(\boldsymbol{z}^{view})}, \tag{8}$$

where $\delta^{(\boldsymbol{z}^{view})}$ denotes a Dirac delta centered at $\boldsymbol{z}^{view}$. Hence, more in general, we could wonder how to process a different statement by simply replacing the Dirac delta $\delta^{(\boldsymbol{z}^{view})}$ with another arbitrary distribution $f_{\boldsymbol{Z}}^{view}$ ("How will behave tomorrow's returns of the S&P 500, if the index return is normally distributed $N(0,1)$?").

To summarize our new problem can be formulated in full generality as follows.

What is the conditional distribution of the market variables, given a **distributional view**, i.e. a distribution $f_{\boldsymbol{Z}}^{view}$ for the factors

$$\boldsymbol{X}|\{\bar{f}_{\boldsymbol{Z}} = f_{\boldsymbol{Z}}^{view}\} \sim ? \tag{9}$$

Following the same rationale, we can further generalize views processing problems as in (9). As a matter of fact, a distributional view can be interpreted as a statement on the features that the updated distribution $\bar{f}_{\boldsymbol{Z}}$ of the view variables has to satisfy

$$\boldsymbol{Z} \sim f_{\boldsymbol{Z}}^{view} \quad \Leftrightarrow \quad \bar{f}_{\boldsymbol{Z}} \in \{f_{\boldsymbol{Z}}^{view}\}. \tag{10}$$

Hence, more in general, we could wonder how to process a different statement by simply replacing the one-element set $\{f_{\boldsymbol{Z}}^{view}\}$ with an arbitrary family of distributions $\mathcal{C}_{\boldsymbol{Z}}$ specified by a collection features or constraints that the view variables have to satisfy ("How will behave tomorrow's returns of the S&P 500, if the *expected* index return is 0%?").

Then the problem can be formulated as follows.

What is the conditional distribution of the market variables, given a **partial view**, i.e. a set of constraints $\mathcal{C}_{\boldsymbol{Z}}$ on the factors

$$\boldsymbol{X}|\{\bar{f}_{\boldsymbol{Z}} \in \mathcal{C}_{\boldsymbol{Z}}\} \sim ? \tag{11}$$

A relevant class of partial views as in (11) are those that can be expressed in terms of equality or inequality expectation conditions

$$\mathcal{C}_{\boldsymbol{Z}} \equiv \{f_{\boldsymbol{Z}} : \quad \mathbb{E}^{f_{\boldsymbol{Z}}}\{\boldsymbol{Z}\} \leqq \boldsymbol{\eta}^{view}\}. \tag{12}$$

where $\boldsymbol{\eta}^{view} \equiv (\eta_1^{view}, \ldots, \eta_{\bar{k}}^{view})'$ are the **features**, i.e. the $\bar{k} \times 1$ vector quantifying the views.

In practice the variables $\boldsymbol{\eta}^{view}$ have a different meaning depending on the application we are pursuing, such as for *prediction* or *stress testing* (what-if analysis) purposes. When we perform prediction, $\boldsymbol{\eta}^{view}$ is a vector of *numbers* that can be computed from current data, say trading signals. When we perform stress testing, $\boldsymbol{\eta}^{view}$ is a vector of *parameters* that we let span in a given range of extreme scenarios ("How will behave tomorrow's returns of the S&P 500, if the tomorrow's expected index return is less than -0.10%?").

The (in)equality conditions (12), though very simple, cover a wide range of practical views, such as on volatilities, correlations, tail behaviors, etc. See also [Meucci, 2008].

General updated distribution

According to the general intuition in (6), the **updated distribution** $\bar{f}_{\boldsymbol{X}}$ is by definition the conditional distribution of the market variables, given the views, which in full generality we can assume to be partial (11)

$$\boldsymbol{X}|\{\bar{f}_{\boldsymbol{Z}} \in \mathcal{C}_{\boldsymbol{Z}}\} \sim \bar{f}_{\boldsymbol{X}}, \tag{13}$$

since both point (7) and distributional (9) views are simpler sub-cases.

In the case of point views as in (7), the updated distribution is defined through the well-known **Bayes' rule**

$$\bar{f}_{\boldsymbol{X}}(\boldsymbol{x}) \equiv \underline{f}_{\boldsymbol{X}}(\boldsymbol{x}|\boldsymbol{z}^{view}) = \frac{\underline{f}_{\boldsymbol{X},\boldsymbol{Z}}(\boldsymbol{x}, \boldsymbol{z}^{view})}{\int_{\mathbb{R}^{\bar{n}}} \underline{f}_{\boldsymbol{X},\boldsymbol{Z}}(\boldsymbol{y}, \boldsymbol{z}^{view}) d\boldsymbol{y}}. \tag{14}$$

In the case of distributional views as in (7), the updated distribution is defined through **conditioning-marginalization**

$$\bar{f}_{\boldsymbol{X}}(\boldsymbol{x}) \equiv \int \underline{f}_{\boldsymbol{X}}(\boldsymbol{x}|\boldsymbol{z}) f^{view}_{\boldsymbol{Z}}(\boldsymbol{z}) d\boldsymbol{z}, \tag{15}$$

which naturally extends the Bayes' rule (14).

In the case of partial views as in (11), we can define a suitable conditioning rule through the **minimum relative entropy principle** (MRE).

More precisely, the updated distribution through MRE, or **MRE updated distribution**, is defined as the one distribution which is the most similar to the base $\underline{f}_{\boldsymbol{X}}$ (1), but at the same time satisfies the views

$$\bar{f}_{\boldsymbol{X}} \equiv \operatorname{argmin}_{f_{\boldsymbol{X}} \in \mathcal{C}_{\boldsymbol{X}}} \mathcal{E}(f_{\boldsymbol{X}} \| \underline{f}_{\boldsymbol{X}}). \tag{16}$$

In (16) $\mathcal{E}$ denotes the **relative entropy**, which is a *pseudo-distance* between distributions

$$\mathcal{E}(f_{\boldsymbol{X}} \| \underline{f}_{\boldsymbol{X}}) \equiv \int_{\mathbb{R}^{\bar{n}}} f_{\boldsymbol{X}}(\boldsymbol{x}) \ln\left(\frac{f_{\boldsymbol{X}}(\boldsymbol{x})}{\underline{f}_{\boldsymbol{X}}(\boldsymbol{x})}\right) d\boldsymbol{x}; \tag{17}$$

while $\mathcal{C}_{\boldsymbol{X}}$ denotes the set of constraints which are implicitly induced by the views[1]

$$\mathcal{C}_{\boldsymbol{X}} \equiv \{f_{\boldsymbol{X}} : \quad f_{\boldsymbol{Z}} \in \mathcal{C}_{\boldsymbol{Z}}\}. \tag{18}$$

Hence, the MRE updated distribution by definition (16) satisfies the constraints (18)

$$\bar{f}_{\boldsymbol{X}} \in \mathcal{C}_{\boldsymbol{X}}. \tag{19}$$

MRE approach naturally extends the conditioning-marginalization (15) and hence the Bayes' rule (14) in turn. See also [Caticha and Giffin, 2006]. Moreover, MRE is a *consistent* conditioning rule, in that it does not perturb the base (1) if it already satisfies the views (18)

$$\underline{f}_{\boldsymbol{X}} \in \mathcal{C}_{\boldsymbol{X}} \quad \Rightarrow \quad \bar{f}_{\boldsymbol{X}} = \underline{f}_{\boldsymbol{X}}. \tag{20}$$

[1] We remember that the view variables are endogenous according to (2).

Exponential family updated distribution

In practice, performing conditioning either via Bayes' rule (14) or conditioning-marginalization (16) or, more in general, via MRE (16) is typically difficult. As a matter of fact, each conditioning techniques requires a large multivariate integration that in general cannot be simply addressed either analytically or numerically. However, according to [Cover and Thomas, 2006], there exists a very special situation where the solution of the MRE problem (16) can be parametrized within a suitable distributional class, see also Figure 1.

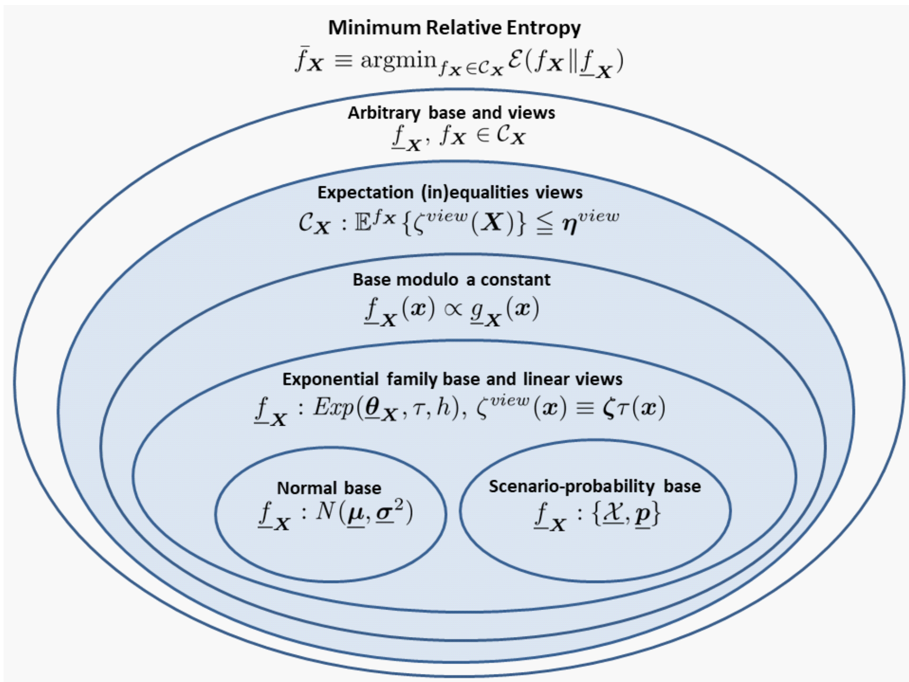

Figure 1: Minimum Relative Entropy problem and sub-cases. We highlight in light blue the conditions under which the MRE updated distribution (16) belongs to the exponential family class (23). In this thesis we investigate special cases under normal base (Chapters 1-2) and scenario-probability base (Chapter 3)

More precisely we have the following theoretical result.

Suppose:

- an arbitrary base distribution (1)

$$\boldsymbol{X} \sim \underline{f}_{\boldsymbol{X}}. \tag{21}$$

- (in)equality views on generalized expectation (12)

$$f_{\boldsymbol{X}} \in \mathcal{C}_{\boldsymbol{X}} : \qquad \mathbb{E}^{f_{\boldsymbol{X}}}\left\{\zeta^{view}(\boldsymbol{X})\right\} \leqq \boldsymbol{\eta}^{view}. \tag{22}$$

Then, the MRE updated distribution (16) belong to the exponential family $Exp(\boldsymbol{\theta}^{view}, \zeta^{view}, \underline{f}_{\boldsymbol{X}})$, i.e.

$$\boldsymbol{X} \sim \bar{f}_{\boldsymbol{X}} : \qquad \bar{f}_{\boldsymbol{X}}(\boldsymbol{x}) \propto \underline{f}_{\boldsymbol{X}}(\boldsymbol{x}) e^{\boldsymbol{\theta}^{view\prime} \zeta^{view}(\boldsymbol{x})}, \tag{23}$$

where the canonical coordinates $\boldsymbol{\theta}^{view} \equiv (\theta_1^{view}, \ldots, \theta_{\bar{k}}^{view})'$ are Lagrange multipliers.

The exponential family (23) is widely flexible for any practical modelling and includes many of the most common distributions, such as the normal, exponential, chi-squared, Wishart, etc.

In particular, an element in the generic exponential family $Exp(\boldsymbol{t}, \zeta^{view}, \underline{f}_{\boldsymbol{X}})$ as in (23) explicitly reads

$$\ln f_{\boldsymbol{t}}(\boldsymbol{x}) = \ln \underline{f}_{\boldsymbol{X}}(\boldsymbol{x}) + \boldsymbol{t}' \zeta^{view}(\boldsymbol{x}) - \psi(\boldsymbol{t}), \tag{24}$$

where ψ denotes the **log-partition function**, which normalizes $f_{\boldsymbol{t}}$ to integrate to one

$$\psi(\boldsymbol{t}) \equiv \ln \int_{\mathbb{R}^{\bar{n}}} e^{\boldsymbol{t}' \zeta^{view}(\boldsymbol{x})} \underline{f}_{\boldsymbol{X}}(\boldsymbol{x})\, d\boldsymbol{x}. \tag{25}$$

Note that the base (1) belongs to the same exponential family class (23), since we have

$$\underline{f}_{\boldsymbol{X}} = f_{\boldsymbol{0}}. \tag{26}$$

Hence, the distributional model $Exp(\boldsymbol{t}, \zeta^{view}, \underline{f}_{\boldsymbol{X}})$ naturally embeds both base $\underline{f}_{\boldsymbol{X}}$ and updated $\bar{f}_{\boldsymbol{X}}$.

Finally, in (24) the optimal Lagrange multipliers $\boldsymbol{\theta}^{view}$ identifying the MRE updated distribution $\bar{f}_{\boldsymbol{X}} \equiv f_{\boldsymbol{\theta}^{view}}$ (23) are the solutions of the following dual Lagrangian problem (see also [Jaakkola, 1999])

$$\boldsymbol{\theta}^{view} \equiv \underset{\boldsymbol{t} \leqq \boldsymbol{0}}{\operatorname{argmin}}\, \mathcal{L}(\boldsymbol{t}; \boldsymbol{\eta}^{view}), \tag{27}$$

where $\mathcal{L}(\boldsymbol{t}; \boldsymbol{\eta}^{view})$ denotes the dual Lagrangian

$$\mathcal{L}(\boldsymbol{t}; \boldsymbol{\eta}^{view}) \equiv \psi(\boldsymbol{t}) - \boldsymbol{t}' \boldsymbol{\eta}^{view}, \tag{28}$$

which is an instance of convex programming, since the log-partition function $\psi(\boldsymbol{t})$ is convex, and as such, it admits a unique solution. As a matter of fact, the gradient of the dual Lagrangian explicitly reads

$$\nabla_{\boldsymbol{t}} \mathcal{L}(\boldsymbol{t}; \boldsymbol{\eta}^{view}) = \mathbb{E}^{f_{\boldsymbol{t}}}\{\zeta^{view}(\boldsymbol{X})\} - \boldsymbol{\eta}^{view}; \tag{29}$$

which shows that at the minimum $\boldsymbol{\theta}^{view}$ for the Lagrangian the views (22) are satisfied; and its Hessian reads

$$\nabla^2_{\boldsymbol{t},\boldsymbol{t}} \mathcal{L}(\boldsymbol{t}; \boldsymbol{\eta}^{view}) = \mathbb{C}v^{f_{\boldsymbol{t}}}\{\zeta^{view}(\boldsymbol{X})\}, \tag{30}$$

see also [Amari and Nagaoka, 2000], [Amari, 2016] and [Schofield, 2007].

Note that the dual Lagrangian problem (27) is an instance of maximum likelihood optimization, as follows because the expected log-likelihood is equivalent to the negative dual Lagrangian (28)

$$\mathbb{E}^{\bar{f}_{\boldsymbol{X}}}\{\ln f_{\boldsymbol{t}}(\boldsymbol{X})\} = \mathbb{E}^{\bar{f}_{\boldsymbol{X}}}\{\ln \underline{f}_{\boldsymbol{X}}(\boldsymbol{X})\} - \mathcal{L}(\boldsymbol{t}; \boldsymbol{\eta}^{view}). \tag{31}$$

However, the dual Lagrangian $\mathcal{L}(\boldsymbol{t}; \boldsymbol{\eta}^{view})$, as well as its derivatives, is *not* analytically tractable in general, and for this reason it is impossible to explicitly solve the optimization (27), unless for a very specific class of base distributions. Refer to Figure 1.

Chapter 1

Analytical solutions for fitting MRE models under normality

1.1 Introduction

In 1990, [Black and Litterman, 1990] (BL in the sequel) provided a mathematical framework for portfolio allocation developed to overcome the problem of computing reasonable estimates of expected returns in the implementations of the "Modern Portfolio Theory", or mean-variance approach, introduced by [Markowitz, 1952]. This model starts with the assumption that the initial, or base, expected returns are set implicitly in terms of CAPM-like *market equilibrium*, and then updated analytically through Bayesian methodology in order to take into account bullish/bearish *views on arbitrary portfolios* and corresponding users's degree of confidence about such views.

The main ingredients behind BL approach are: i) the *normality* underlying the base reference model; ii) the *linearity* of the functions specifying the view variables; and iii) the format of the views as *equality* (distributional (10)) *statements*, where the uncertainty around the view-implied expectations is the same as the one induced by the base distribution.

According to this setup, the BL method generalizes also the standard scenario analysis a-la RiskMetrics (see [RiskMetrics, 1996], [Mina and Xiao, 2001]), which processes equality statements, or point views (7), on the future realizations of the view variables (special case of no uncertainty on the expected view variables).

In the following years many other advanced implementations for view-processing have been developed, see for instance [Pezier, 2007], [Almgren and Chriss, 2006]. In particular, under the normal assumption, [Qian and Gorman, 2001] proposed a Bayesian-like approach allowing to process any arbitrary normal view (10), generalizing BL in turn.

Accordingly, [Meucci, 2008] showed how the above normal implementations were all instances of the MRE principle (see also [Meucci, 2010]) and generalized the formulations by [Black and Litterman, 1990], [Qian and Gorman, 2001] to handle partial views (11) on arbitrary linear combinations of expectations and covariances.

Here we enhance and generalize results from [Meucci, 2008]. More precisely, we present how to derive the canonical formulation of the MRE updated distribution within the exponential-family class and according to the parametric implementations of the MRE. In particular, we compute the optimal Lagrange multipliers driving the MRE updated distribution under views on linear combinations of the first two moments of the market variables.

This will provide a background for covering more general (in)equality views, such as on ranking, as we discuss in Chapter 2, and a benchmark for sample-based implementations of MRE under non-normal markets, as we discuss in Chapter 3.

The remainder of this chapter is organized as follows.

In Section 1.2 we introduce the MRE theoretical framework for exponential family base distributions and equality views on moments conditions. In Section 1.3 we present an analytical solutions for linear views on first moments, which include [Black and Litterman, 1990] and [Mina and Xiao, 2001] as special case. In Section 1.4 we discuss the case of linear views on (non-central) second moments and show how to address the MRE solution via numerical recursions or analytically under views second central moments. In Section 1.5 we illustrate the more general solutions under joint views on the first two (non-central) moments, which are addressed similar to Section 1.4 and extend formulations by [Qian and Gorman, 2001] and [Meucci, 2008] in turn. Finally, in Section 1.6 we list the main contributions.

Fully documented code is available on GitHub.

1.2 The model

Following the theoretical framework (4), here we address analytically the MRE problem (16) under the assumption that the base distribution (1) belongs to a specific exponential family class, see Figure 1. Refer to [Amari and Nagaoka, 2000] and [Amari, 2016] for more details.

More precisely, let us consider the following setup.

Suppose:
- a base distribution (1) which belongs to an exponential family class $Exp(\underline{\boldsymbol{\theta}}_{\boldsymbol{X}}, \tau, h)$

$$\boldsymbol{X} \sim \underline{f}_{\boldsymbol{X}} : \qquad \underline{f}_{\boldsymbol{X}}(\boldsymbol{x}) \propto h(\boldsymbol{x}) e^{\underline{\boldsymbol{\theta}}'_{\boldsymbol{X}} \tau(\boldsymbol{x})}, \tag{1.1}$$

where $\underline{\boldsymbol{\theta}}_{\boldsymbol{X}} \equiv (\underline{\theta}_{\boldsymbol{X};1}, \ldots, \underline{\theta}_{\boldsymbol{X};\bar{l}})'$ is a vector of base canonical parameters; $\tau(\boldsymbol{x}) \equiv (\tau_1(\boldsymbol{x}), \ldots, \tau_{\bar{l}}(\boldsymbol{x}))'$ are sufficient statistics and $h(\boldsymbol{x}) > 0$ is a reference measure;
- equality views on generalized expectation as in (22)

$$f_{\boldsymbol{X}} \in \mathcal{C}_{\boldsymbol{X}} : \qquad \mathbb{E}^{f_{\boldsymbol{X}}} \left\{ \zeta^{view}(\boldsymbol{X}) \right\} = \boldsymbol{\eta}^{view}, \tag{1.2}$$

where the view function ζ^{view} is *linear* in the sufficient statistics τ, or

$$\zeta^{view}(\boldsymbol{x}) \equiv \boldsymbol{\zeta} \tau(\boldsymbol{x}), \tag{1.3}$$

and where $\boldsymbol{\zeta}$ is a $\bar{k} \times \bar{l}$ matrix.
Then, the MRE updated distribution (23) must belong to the *same* exponential family class as the base [A.1.1]

$$\boldsymbol{X} \sim \bar{f}_{\boldsymbol{X}} : \qquad \bar{f}_{\boldsymbol{X}}(\boldsymbol{x}) \propto h(\boldsymbol{x}) e^{\bar{\boldsymbol{\theta}}'_{\boldsymbol{X}} \tau(\boldsymbol{x})}, \tag{1.4}$$

where $\bar{\boldsymbol{\theta}}_{\boldsymbol{X}} \equiv (\bar{\theta}_{\boldsymbol{X};1}, \ldots, \bar{\theta}_{\boldsymbol{X};\bar{l}})'$ is a vector of updated canonical parameters

$$\bar{\boldsymbol{\theta}}_{\boldsymbol{X}} = \underline{\boldsymbol{\theta}}_{\boldsymbol{X}} + \boldsymbol{\zeta}' \boldsymbol{\theta}^{view}, \tag{1.5}$$

and $\boldsymbol{\theta}^{view} \equiv (\theta_1^{view}, \ldots, \theta_{\bar{k}}^{view})'$ is a $\bar{k} \times 1$ vector of optimal Lagrange multipliers (27), that needs to be computed.

In particular, here we focus on:
- a multivariate normal base distribution (1)

$$\underline{f}_{\boldsymbol{X}} \quad \Leftrightarrow \quad N(\underline{\boldsymbol{\mu}}_{\boldsymbol{X}}, \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}), \tag{1.6}$$

where $\underline{\boldsymbol{\mu}}_{\boldsymbol{X}}$ is an $\bar{n} \times 1$ vector and $\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}$ is an $\bar{n} \times \bar{n}$ symmetric and positive definite matrix;
- views on the first two moments of the market variables

$$f_{\boldsymbol{X}} \in \mathcal{C}_{\boldsymbol{X}} : \qquad \begin{cases} \mathbb{E}^{f_{\boldsymbol{X}}}\{\boldsymbol{\zeta}_\mu \boldsymbol{X}\} = \boldsymbol{\eta}_\mu^{view} \\ \mathbb{E}^{f_{\boldsymbol{X}}}\{\boldsymbol{\zeta}_{\sigma,\sigma} vec(\boldsymbol{X}\boldsymbol{X}')\} = vec(\boldsymbol{\eta}_{\sigma,\sigma}^{view}), \end{cases} \tag{1.7}$$

where $\boldsymbol{\eta}_\mu^{view}$ is a $\bar{k}_\mu \times 1$ vector and $\boldsymbol{\zeta}_\mu$ is a $\bar{k}_\mu \times \bar{n}$ matrix; $\boldsymbol{\eta}_{\sigma,\sigma}^{view}$ is a $\bar{k}_\sigma \times \bar{k}_\sigma$ symmetric matrix (without loss of generality) and $\boldsymbol{\zeta}_{\sigma,\sigma}$ is a $\bar{k}_\sigma^2 \times \bar{n}^2$ matrix. Then the effective number of views is $\bar{k}_\mu + \bar{k}_\sigma(\bar{k}_\sigma + 1)/2$. Equivalently, we can always assume $\boldsymbol{\zeta}_{\sigma,\sigma}$ to be arranged as follows

$$\boldsymbol{\zeta}_{\sigma,\sigma} \equiv \boldsymbol{\zeta}_\sigma \otimes \boldsymbol{\zeta}_\sigma, \tag{1.8}$$

for a suitable $\bar{k}_\sigma \times \bar{n}$ matrix $\boldsymbol{\zeta}_\sigma$, so that the views (1.7) becomes

$$f_{\boldsymbol{X}} \in \mathcal{C}_{\boldsymbol{X}} : \qquad \begin{cases} \mathbb{E}^{f_{\boldsymbol{X}}}\{\boldsymbol{\zeta}_\mu \boldsymbol{X}\} = \boldsymbol{\eta}_\mu^{view} \\ \mathbb{E}^{f_{\boldsymbol{X}}}\{\boldsymbol{\zeta}_\sigma \boldsymbol{X}\boldsymbol{X}' \boldsymbol{\zeta}_\sigma'\} = \boldsymbol{\eta}_{\sigma,\sigma}^{view}, \end{cases} \tag{1.9}$$

as follows from the properties of the Kronecker product [Magnus and Neudecker, 1979].

The assumption of a normal base distribution with views on the first two moments (1.6)-(1.9) is a special case of (1.1)-(1.2). Indeed, the normal distribution is a special exponential family distribution as in (1.1)

$$N(\underline{\boldsymbol{\mu}}_{\boldsymbol{X}}, \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}) \quad \Leftrightarrow \quad Exp(\underline{\boldsymbol{\theta}}_{\boldsymbol{X}}, \tau, h), \tag{1.10}$$

where the $\bar{n} \times 1$ vector canonical coordinates $\underline{\boldsymbol{\theta}}_{\boldsymbol{X}}$ read

$$\underline{\boldsymbol{\theta}}_{\boldsymbol{X}} \equiv \begin{pmatrix} \underline{\boldsymbol{\theta}}_{\boldsymbol{X};\mu} \\ vec(\underline{\boldsymbol{\theta}}_{\boldsymbol{X};\sigma,\sigma}) \end{pmatrix} = \begin{pmatrix} (\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}})^{-1} \underline{\boldsymbol{\mu}}_{\boldsymbol{X}} \\ -\frac{1}{2} vec((\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}})^{-1}) \end{pmatrix}; \tag{1.11}$$

the sufficient statistics read

$$\tau(\boldsymbol{x}) \equiv \left(\begin{array}{c} \tau_{\mu}(\boldsymbol{x}) \\ \tau_{\sigma,\sigma}(\boldsymbol{x}) \end{array} \right) \equiv \left(\begin{array}{c} \boldsymbol{x} \\ vec(\boldsymbol{x}\boldsymbol{x}') \end{array} \right); \tag{1.12}$$

and the reference measure reads $h(\boldsymbol{x}) \equiv (2\pi)^{-\bar{n}/2}$.

Furthermore, the $\bar{k} \equiv \bar{k}_{\mu} + \bar{k}_{\sigma}^2$ views on the first two moments (1.9) are special views on generalized expectation as in (1.2), where:
- the $\bar{k} \times (\bar{n} + \bar{n}^2)$ matrix $\boldsymbol{\zeta}$ in (1.3) reads as follows

$$\boldsymbol{\zeta} \equiv \begin{pmatrix} \boldsymbol{\zeta}_{\mu} & \boldsymbol{0}_{\bar{k}_{\mu} \times \bar{n}^2} \\ \boldsymbol{0}_{\bar{k}_{\sigma}^2 \times \bar{n}} & \boldsymbol{\zeta}_{\sigma,\sigma} \end{pmatrix}; \tag{1.13}$$

- the $\bar{k} \times 1$ vector $\boldsymbol{\eta}^{view}$ in (1.2) reads as follows

$$\boldsymbol{\eta}^{view} \equiv \left(\begin{array}{c} \boldsymbol{\eta}_{\mu}^{view} \\ vec(\boldsymbol{\eta}_{\sigma,\sigma}^{view}) \end{array} \right). \tag{1.14}$$

Then, according to the general framework (1.4), the ensuing updated distribution (23) must be normal in turn

$$\bar{f}_{\boldsymbol{X}} \quad \Leftrightarrow \quad Exp(\bar{\boldsymbol{\theta}}_{\boldsymbol{X}}, \tau, h) \quad \Leftrightarrow N(\bar{\boldsymbol{\mu}}_{\boldsymbol{X}}, \bar{\boldsymbol{\sigma}}_{\boldsymbol{X}}^2), \tag{1.15}$$

see also [Cover and Thomas, 2006]. In (1.15) the $\bar{n} \times 1$ vector updated canonical coordinates $\bar{\boldsymbol{\theta}}_{\boldsymbol{X}}$ read as in (1.5) and where $\boldsymbol{\theta}^{view}$ is the $\bar{k} \times 1$ vector of optimal Lagrange multipliers (27), which we arrange as follows

$$\boldsymbol{\theta}^{view} \equiv \left(\begin{array}{c} \boldsymbol{\theta}_{\mu}^{view} \\ vec(\boldsymbol{\theta}_{\sigma,\sigma}^{view}) \end{array} \right). \tag{1.16}$$

Finally, the updated expectation in (1.15) follows from the updated canonical coordinates $\bar{\boldsymbol{\theta}}_{\boldsymbol{X}}$ (1.5) [A.1.2]

$$\bar{\boldsymbol{\mu}}_{\boldsymbol{X}} = -\frac{1}{2}(\bar{\boldsymbol{\theta}}_{\boldsymbol{X};\sigma,\sigma})^{-1}\bar{\boldsymbol{\theta}}_{\boldsymbol{X};\mu}, \tag{1.17}$$

and similar for the updated covariance in (1.15)

$$\bar{\boldsymbol{\sigma}}_{\boldsymbol{X}}^2 = -\frac{1}{2}(\bar{\boldsymbol{\theta}}_{\boldsymbol{X};\sigma,\sigma})^{-1}. \tag{1.18}$$

Refer to [A.1.3] for a more explicit expression of (1.17)-(1.18).

Now our goal is to compute the optimal Lagrange multipliers $\boldsymbol{\theta}^{view}$. To this purpose we proceed step by step.

1.3 Views on first moments

Let us suppose the case of only "views on portfolios" a-la Black-Litterman, i.e. equality views (18) on linear combinations of the first moments

$$f_{\boldsymbol{X}} \in \mathcal{C}_{\boldsymbol{X}}: \qquad \mathbb{E}^{f_{\boldsymbol{X}}}\{\boldsymbol{\zeta}_\mu \boldsymbol{X}\} = \boldsymbol{\eta}_\mu^{view}, \tag{1.19}$$

where $\boldsymbol{\zeta}_\mu$ is a $\bar{k} \times \bar{n}$ full rank matrix. Then the optimal Lagrange multipliers $\boldsymbol{\theta}^{view} = \boldsymbol{\theta}_\mu^{view}$ in (1.16) can be computed analytically and read [A.1.4]

$$\boldsymbol{\theta}_\mu^{view} = (\boldsymbol{\zeta}_\mu \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} \boldsymbol{\zeta}_\mu')^{-1}(\boldsymbol{\eta}_\mu^{view} - \boldsymbol{\zeta}_\mu \underline{\boldsymbol{\mu}}_{\boldsymbol{X}}). \tag{1.20}$$

From (1.5) we obtain explicitly the updated expectation (1.17) [A.1.4]

$$\bar{\boldsymbol{\mu}}_{\boldsymbol{X}} = \underline{\boldsymbol{\mu}}_{\boldsymbol{X}} + \boldsymbol{\zeta}_\mu^{\dagger\prime}(\boldsymbol{\eta}_\mu^{view} - \boldsymbol{\zeta}_\mu \underline{\boldsymbol{\mu}}_{\boldsymbol{X}}), \tag{1.21}$$

where $\boldsymbol{\zeta}_\mu^\dagger$ is a $\bar{k} \times \bar{n}$ pseudo-inverse matrix for $\boldsymbol{\zeta}_\mu'$

$$\boldsymbol{\zeta}_\mu^\dagger \equiv (\boldsymbol{\zeta}_\mu \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} \boldsymbol{\zeta}_\mu')^{-1} \boldsymbol{\zeta}_\mu \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}; \tag{1.22}$$

and the updated covariance (1.18)

$$\bar{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} = \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}. \tag{1.23}$$

To summarize

Optimal Lagr. mult.	
$\boldsymbol{\theta}_\mu^{view} = (\boldsymbol{\zeta}_\mu \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} \boldsymbol{\zeta}_\mu')^{-1}(\boldsymbol{\eta}_\mu^{view} - \boldsymbol{\zeta}_\mu \underline{\boldsymbol{\mu}}_{\boldsymbol{X}})$	$\boldsymbol{\theta}_{\sigma,\sigma}^{view} = \varnothing$
Updated exp. and cov.	
$\bar{\boldsymbol{\mu}}_{\boldsymbol{X}} = \underline{\boldsymbol{\mu}}_{\boldsymbol{X}} + \boldsymbol{\zeta}_\mu^{\dagger\prime}(\boldsymbol{\eta}_\mu^{view} - \boldsymbol{\zeta}_\mu \underline{\boldsymbol{\mu}}_{\boldsymbol{X}})$	$\bar{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} = \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}$

Table 1.1: Views on first moments: MRE solutions

Note how the pseudo-inverse $\boldsymbol{\zeta}_\mu^\dagger$ (1.22) can be equivalently weighted using either i) the base covariance $\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}$, or ii) the updated covariance $\bar{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}$, since in this case they both coincide (1.23). As we will show later, this feature will persist even in more general situations, where the identity (1.23) does *not* hold.

Finally, we observe that the updated expectation (1.21) can be interpreted as a projection. Indeed, consider the $(\bar{n} - \bar{k})$-hyperplane defined as follows

$$\mathcal{S}_{\boldsymbol{\zeta}_\mu} \equiv \{\boldsymbol{\mu} \in \mathbb{R}^{\bar{n}} \text{ such that } \boldsymbol{\zeta}_\mu \boldsymbol{\mu} = \boldsymbol{0}\}; \tag{1.24}$$

and the projection operator of an $\bar{n} \times 1$ vector $\boldsymbol{x}$ onto $\mathcal{S}_{\boldsymbol{\zeta}_\mu}$

$$\mathfrak{P}_\mu[\boldsymbol{x}] \equiv (\mathbb{I}_{\bar{n}} - \boldsymbol{\zeta}_\mu^{\dagger\prime} \boldsymbol{\zeta}_\mu)\boldsymbol{x}, \tag{1.25}$$

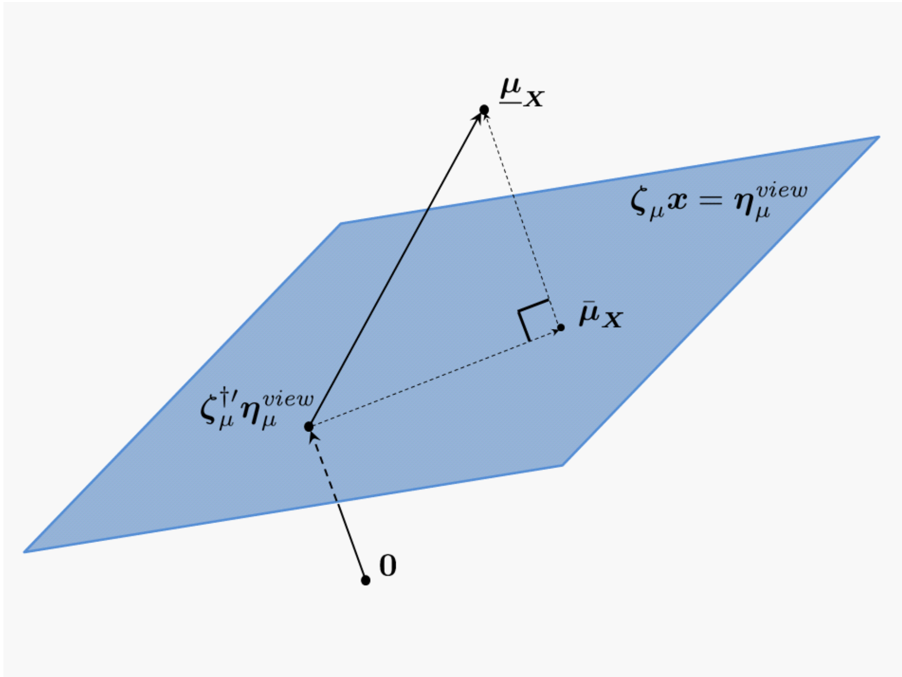

Figure 1.1: Updated expectation as orthogonal projection

which is orthogonal with respect to the inner product $\langle \boldsymbol{x}, \boldsymbol{y} \rangle \equiv \boldsymbol{x}'(\boldsymbol{\sigma}^2_{\boldsymbol{X}})^{-1}\boldsymbol{y}$ induced by the inverse base, or updated, covariance (1.23).

Then the updated expectation (1.21) reads [A.1.5]

$$\bar{\boldsymbol{\mu}}_{\boldsymbol{X}} = \boldsymbol{\zeta}_\mu^{\dagger\prime} \boldsymbol{\eta}_\mu^{view} + \mathfrak{P}_\mu[\underline{\boldsymbol{\mu}}_{\boldsymbol{X}} - \boldsymbol{\zeta}_\mu^{\dagger\prime} \boldsymbol{\eta}_\mu^{view}], \tag{1.26}$$

see also Figure 1.1.

Not surprisingly, when we consider views on expectations as in (1.19), the only expectation is updated in order to satisfy the views, while the covariance is the same as the base counterpart.

This is consistent with the Black-Litterman solution [Black and Litterman, 1990], which is a special case of [Meucci, 2010].

Example 1.1. Consider $\bar{n} \equiv 3$ market variables $\boldsymbol{X} \equiv (X_1, X_2, X_3)'$ with joint normal base distribution

$$\boldsymbol{X} \sim N(\underline{\boldsymbol{\mu}}_{\boldsymbol{X}}, \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}), \tag{1.27}$$

where

$$\underline{\boldsymbol{\mu}}_{\boldsymbol{X}} \equiv \begin{pmatrix} 0.26 \\ 0.29 \\ 0.33 \end{pmatrix}, \quad \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} \equiv \begin{pmatrix} 0.18 & 0.11 & 0.13 \\ 0.11 & 0.23 & 0.16 \\ 0.13 & 0.16 & 0.23 \end{pmatrix}, \tag{1.28}$$

and assume the true updated distribution $\boldsymbol{X} \sim Exp(\boldsymbol{\theta}_{\mu}^{view*}, \zeta^{view}, \underline{f}_{\boldsymbol{X}})$ (1.15) to be steered by the following Lagrange multipliers

$$\boldsymbol{\theta}_{\mu}^{view*} = \left(\begin{smallmatrix} 5.71 \\ 0.38 \end{smallmatrix}\right); \tag{1.29}$$

with $\bar{k} = 2$ view variables specified by $\zeta^{view}(\boldsymbol{x}) \equiv \boldsymbol{\zeta}_{\mu}\boldsymbol{x}$ (1.3), where

$$\boldsymbol{\zeta}_{\mu} \equiv \left(\begin{smallmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{smallmatrix}\right). \tag{1.30}$$

Then, if we input the true $\bar{k} \times 1$ vector of features (1.19) implied by the true updated distribution

$$\boldsymbol{\eta}_{\mu}^{view} \equiv \boldsymbol{\zeta}_{\mu}\bar{\boldsymbol{\mu}}_{\boldsymbol{X}}^{*} = \left(\begin{smallmatrix} 1.02 \\ -0.50 \end{smallmatrix}\right), \tag{1.31}$$

in the formulation of the optimal Lagrange multipliers $\boldsymbol{\theta}_{\mu}^{view}$ (1.20), we obtain back the true counterpart

$$||\boldsymbol{\theta}_{\mu}^{view*} - \boldsymbol{\theta}_{\mu}^{view}|| = 2.79 \times 10^{-15}. \tag{1.32}$$

Similar results follow for expectation $\bar{\boldsymbol{\mu}}_{\boldsymbol{X}}$ (1.21) and covariance $\bar{\boldsymbol{\sigma}}_{\boldsymbol{X}}^{2}$ (1.23).

1.4 Views on second moments

Let us suppose the case of only equality views (18) on linear combinations of the second non-central moments

$$f_{\boldsymbol{X}} \in \mathcal{C}_{\boldsymbol{X}}: \qquad \mathbb{E}^{f_{\boldsymbol{X}}}\{\boldsymbol{\zeta}_{\sigma}\boldsymbol{X}\boldsymbol{X}'\boldsymbol{\zeta}_{\sigma}'\} = \boldsymbol{\eta}_{\sigma,\sigma}^{view}, \tag{1.33}$$

where $\boldsymbol{\zeta}_{\sigma}$ is a $\bar{k} \times \bar{n}$ full rank matrix.

Then, it turns out that the optimal Lagrange multipliers $\boldsymbol{\theta}^{view} = \boldsymbol{\theta}_{\sigma,\sigma}^{view}$ in (1.16) are defined implicitly in terms of the updated expectation implied by the view variables (2)

$$\boldsymbol{\eta}_{\sigma}^{view} \equiv \mathbb{E}^{\bar{f}_{\boldsymbol{X}}}\{\boldsymbol{\zeta}_{\sigma}\boldsymbol{X}\} = \boldsymbol{\zeta}_{\sigma}\bar{\boldsymbol{\mu}}_{\boldsymbol{X}}, \tag{1.34}$$

which is *not* known a priori.

More precisely, if we define the $\bar{k} \times \bar{k}$ matrix-variate function with respect to $\boldsymbol{\eta}_{\sigma}^{view}$

$$\begin{aligned} \sigma^{2view}(\boldsymbol{\eta}_{\sigma}^{view}) &\equiv \boldsymbol{\eta}_{\sigma,\sigma}^{view} - \boldsymbol{\eta}_{\sigma}^{view}\boldsymbol{\eta}_{\sigma}^{view\prime} \\ &= \boldsymbol{\zeta}_{\sigma}\bar{\boldsymbol{\sigma}}_{\boldsymbol{X}}^{2}\boldsymbol{\zeta}_{\sigma}', \end{aligned} \tag{1.35}$$

then we have [A.1.6]

$$\boldsymbol{\theta}_{\sigma,\sigma}^{view} = \frac{1}{2}((\boldsymbol{\zeta}_{\sigma}\underline{\boldsymbol{\sigma}}_{\boldsymbol{X}}^{2}\boldsymbol{\zeta}_{\sigma}')^{-1} - (\sigma^{2view}(\boldsymbol{\eta}_{\sigma}^{view}))^{-1}). \tag{1.36}$$

Similar to the above, from the updated canonical coordinates (1.5), we can deduce also the updated expectation (1.17) [A.1.6]

$$\bar{\boldsymbol{\mu}}_{\boldsymbol{X}} = \underline{\boldsymbol{\mu}}_{\boldsymbol{X}} + \boldsymbol{\zeta}_{\sigma}^{\dagger\prime}(\boldsymbol{\eta}_{\sigma}^{view} - \boldsymbol{\zeta}_{\sigma}\underline{\boldsymbol{\mu}}_{\boldsymbol{X}}), \tag{1.37}$$

and the updated covariance (1.17)

$$\bar{\boldsymbol{\sigma}}_{\boldsymbol{X}}^{2} = \underline{\boldsymbol{\sigma}}_{\boldsymbol{X}}^{2} + \boldsymbol{\zeta}_{\sigma}^{\dagger\prime}(\sigma^{2view}(\boldsymbol{\eta}_{\sigma}^{view}) - \boldsymbol{\zeta}_{\sigma}\underline{\boldsymbol{\sigma}}_{\boldsymbol{X}}^{2}\boldsymbol{\zeta}_{\sigma}^{\prime})\boldsymbol{\zeta}_{\sigma}^{\dagger}, \tag{1.38}$$

where $\boldsymbol{\zeta}_{\sigma}^{\dagger}$ is a $\bar{k} \times \bar{n}$ pseudo-inverse matrix for $\boldsymbol{\zeta}_{\sigma}^{\prime}$ as in (1.22).

Also, it turns out that the pseudo-inverse $\boldsymbol{\zeta}_{\sigma}^{\dagger}$ can be equivalently weighted using either i) the base covariance $\underline{\boldsymbol{\sigma}}_{\boldsymbol{X}}^{2}$, or ii) the updated covariance $\bar{\boldsymbol{\sigma}}_{\boldsymbol{X}}^{2}$, even though they do not coincide [A.1.7]

$$\boldsymbol{\zeta}_{\sigma}^{\dagger} \equiv (\boldsymbol{\zeta}_{\sigma}\underline{\boldsymbol{\sigma}}_{\boldsymbol{X}}^{2}\boldsymbol{\zeta}_{\sigma}^{\prime})^{-1}\boldsymbol{\zeta}_{\sigma}\underline{\boldsymbol{\sigma}}_{\boldsymbol{X}}^{2} = (\boldsymbol{\zeta}_{\sigma}\bar{\boldsymbol{\sigma}}_{\boldsymbol{X}}^{2}\boldsymbol{\zeta}_{\sigma}^{\prime})^{-1}\boldsymbol{\zeta}_{\sigma}\bar{\boldsymbol{\sigma}}_{\boldsymbol{X}}^{2}. \tag{1.39}$$

In conclusion, differently from the case of views on first moments in Section 1.3, when we consider views on second moments as in (1.33), *not* only the covariance, *but* also the expectation is updated.

1.4.1 Numerical solution via recursion

We can address the MRE solutions (1.36)-(1.37)-(1.38) numerically through a simple fixed-point recursion in the view-implied expectation (1.34) as follows from (1.37) [A.1.6]

$$\boldsymbol{\eta}_{\sigma}^{view} = g(\boldsymbol{\eta}_{\sigma}^{view}) \equiv (\boldsymbol{\eta}_{\sigma,\sigma}^{view} - \boldsymbol{\eta}_{\sigma}^{view}\boldsymbol{\eta}_{\sigma}^{view\prime})(\boldsymbol{\zeta}_{\sigma}\underline{\boldsymbol{\sigma}}_{\boldsymbol{X}}^{2}\boldsymbol{\zeta}_{\sigma}^{\prime})^{-1}\boldsymbol{\zeta}_{\sigma}\underline{\boldsymbol{\mu}}_{\boldsymbol{X}}. \tag{1.40}$$

Then, the routine can be set up as follows.

$(\boldsymbol{\theta}_{\sigma,\sigma}^{view}, \bar{\boldsymbol{\mu}}_{\boldsymbol{X}}, \bar{\boldsymbol{\sigma}}_{\boldsymbol{X}}^{2}) \leftarrow Fit.MRE.Second.Moments.N(\boldsymbol{\zeta}_{\sigma}, \boldsymbol{\eta}_{\sigma,\sigma}^{view}, \underline{\boldsymbol{\mu}}_{\boldsymbol{X}}, \underline{\boldsymbol{\sigma}}_{\boldsymbol{X}}^{2})$	
0. Initialize	$\boldsymbol{\eta}_{\sigma}^{view} \leftarrow \boldsymbol{\zeta}_{\sigma}\underline{\boldsymbol{\mu}}_{\boldsymbol{X}}$
1. Update feat.	$\boldsymbol{\eta}_{\sigma}^{view} \leftarrow g(\boldsymbol{\eta}_{\sigma}^{view})$ (1.40)
2. If convergence, output $(\boldsymbol{\theta}_{\sigma,\sigma}^{view}, \bar{\boldsymbol{\mu}}_{\boldsymbol{X}}, \bar{\boldsymbol{\sigma}}_{\boldsymbol{X}}^{2})$ (1.36)-(1.37)-(1.40); else go to 1	

Table 1.2: Iterative routine for optimal Lagrange multipliers under views on second non-central moments

If the true stationary point $\boldsymbol{\eta}_{\sigma}^{view}$ is an *attractive* fixed point for (1.40), the convergence in the above routine occurs when the relative norm between two subsequent updates $\boldsymbol{\eta}_{\sigma}^{view(i)}$ and $\boldsymbol{\eta}_{\sigma}^{view(i+1)}$ is smaller than a required threshold. See [Coxeter, 1998] for more details.

Example 1.2. Consider the same base model as in Example 1.1 and assume the true updated distribution $\boldsymbol{X} \sim Exp(\boldsymbol{\theta}^{view*}_{\sigma,\sigma}, \zeta^{view}, \underline{f}_{\boldsymbol{X}})$ (1.15) to be steered by the following Lagrange multipliers

$$\boldsymbol{\theta}^{view*}_{\sigma,\sigma} = \begin{pmatrix} -1.83 & -2.82 \\ -2.82 & -3.13 \end{pmatrix}; \tag{1.41}$$

with $\bar{k}^2 = 4$ view variables specified by $\zeta^{view}(\boldsymbol{x}) \equiv (\boldsymbol{\zeta}_\sigma \otimes \boldsymbol{\zeta}_\sigma)vec(\boldsymbol{x}\boldsymbol{x}')$ (1.3), where

$$\boldsymbol{\zeta}_\sigma \equiv \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \end{pmatrix}. \tag{1.42}$$

Then, if we input the true $\bar{k} \times \bar{k}$ vector of features (1.19) implied by the true updated distribution

$$\boldsymbol{\eta}^{view}_{\sigma,\sigma} \equiv \boldsymbol{\zeta}_\sigma(\bar{\boldsymbol{\sigma}}^{2*} + \bar{\boldsymbol{\mu}}^*_{\boldsymbol{X}}\bar{\boldsymbol{\mu}}^{*\prime}_{\boldsymbol{X}})\boldsymbol{\zeta}'_\sigma = \begin{pmatrix} 0.19 & -0.19 \\ -0.19 & 0.32 \end{pmatrix}, \tag{1.43}$$

in the recursion in Table 1.2, we obtain back from the optimal Lagrange multipliers $\boldsymbol{\theta}^{view}_{\sigma,\sigma}$ (1.36) the true counterpart

$$||\boldsymbol{\theta}^{view*}_{\sigma,\sigma} - \boldsymbol{\theta}^{view}_{\sigma,\sigma}|| = 1.55 \times 10^{-8}. \tag{1.44}$$

Similar results follow for expectation $\bar{\boldsymbol{\mu}}_{\boldsymbol{X}}$ (1.37) and covariance $\bar{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}$ (1.40).

1.4.2 Analytical solution under special assumptions

We can address the MRE solutions (1.36)-(1.37)-(1.38) analytically under special assumptions on the base distribution and views.

Zero base expectation

In the special case where the base expectation is null, or

$$\underline{\boldsymbol{\mu}}_{\boldsymbol{X}} \equiv \mathbf{0}, \tag{1.45}$$

no recursion is needed. Indeed, the view-implied expectation (1.40) becomes null

$$\boldsymbol{\eta}^{view}_{\sigma} \equiv \mathbb{E}^{\bar{f}_{\boldsymbol{X}}}\{\boldsymbol{\zeta}_\sigma \boldsymbol{X}\} = \mathbf{0}, \tag{1.46}$$

and then both Lagrange multipliers $\boldsymbol{\theta}^{view}_{\sigma,\sigma}$ (1.36), updated expectation $\bar{\boldsymbol{\mu}}_{\boldsymbol{X}}$ (1.37) and covariance $\bar{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}$ (1.38) follow in turn, as summarized in the table below.

Optimal Lagr. mult.	
$\boldsymbol{\theta}_{\mu}^{view} = \varnothing$	$\boldsymbol{\theta}_{\sigma,\sigma}^{view} = \frac{1}{2}((\boldsymbol{\zeta}_{\sigma}\underline{\boldsymbol{\sigma}}_{\boldsymbol{X}}^{2}\boldsymbol{\zeta}_{\sigma}')^{-1} - (\boldsymbol{\eta}_{\sigma,\sigma}^{view})^{-1})$
Updated exp. and cov.	
$\bar{\boldsymbol{\mu}}_{\boldsymbol{X}} = \mathbf{0}$	$\bar{\boldsymbol{\sigma}}_{\boldsymbol{X}}^{2} = \underline{\boldsymbol{\sigma}}_{\boldsymbol{X}}^{2} + \boldsymbol{\zeta}_{\sigma}^{\dagger\prime}(\boldsymbol{\eta}_{\sigma,\sigma}^{view} - \boldsymbol{\zeta}_{\sigma}\underline{\boldsymbol{\sigma}}_{\boldsymbol{X}}^{2}\boldsymbol{\zeta}_{\sigma}')\boldsymbol{\zeta}_{\sigma}^{\dagger}$

Table 1.3: Views on second moments: MRE solutions under null base expectation

Example 1.3. Consider the same base model as in Example 1.1, but with zero base expectation

$$\underline{\boldsymbol{\mu}}_{\boldsymbol{X}} \equiv \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \tag{1.47}$$

and assume the same true updated distribution $\boldsymbol{X} \sim Exp(\boldsymbol{\theta}_{\sigma,\sigma}^{view*}, \zeta^{view}, \underline{f}_{\boldsymbol{X}})$ as specified in Example 1.2. Then, also the true updated expectation is null

$$\bar{\boldsymbol{\mu}}_{\boldsymbol{X}}^{*} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \tag{1.48}$$

and if we input the true $\bar{k}\times\bar{k}$ vector of features (1.19) implied by the true updated distribution

$$\boldsymbol{\eta}_{\sigma,\sigma}^{view} \equiv \boldsymbol{\zeta}_{\sigma}\bar{\boldsymbol{\sigma}}^{2*}\boldsymbol{\zeta}_{\sigma}' = \begin{pmatrix} 0.19 & -0.17 \\ -0.17 & 0.29 \end{pmatrix}, \tag{1.49}$$

in the formulation of the optimal Lagrange multipliers $\boldsymbol{\theta}_{\sigma,\sigma}^{view}$ in Table 1.3, we obtain back the true counterpart

$$||\boldsymbol{\theta}_{\sigma,\sigma}^{view*} - \boldsymbol{\theta}_{\sigma,\sigma}^{view}|| = 9.93 \times 10^{-16}. \tag{1.50}$$

Similar results follow for the covariance $\bar{\boldsymbol{\sigma}}_{\boldsymbol{X}}^{2}$ in Table 1.3.

Views on covariances

Suppose the case of joint equality views (18) on linear combinations of *covariances*

$$f_{\boldsymbol{X}} \in \mathcal{C}_{\boldsymbol{X}}: \qquad \mathbb{C}v^{f_{\boldsymbol{X}}}\{\boldsymbol{\zeta}_{\sigma}\boldsymbol{X}\} = \boldsymbol{\sigma}^{2view}. \tag{1.51}$$

Then the views can be easily re-written as linear combinations of second non-central moments (1.33)

$$f_{\boldsymbol{X}} \in \mathcal{C}_{\boldsymbol{X}}: \qquad \mathbb{E}^{f_{\boldsymbol{X}}}\{\boldsymbol{\zeta}_{\sigma}\boldsymbol{X}\boldsymbol{X}'\boldsymbol{\zeta}_{\sigma}\} = \boldsymbol{\eta}_{\sigma,\sigma}^{view}, \tag{1.52}$$

where the $\bar{k}\times\bar{k}$ matrix of features $\boldsymbol{\eta}_{\sigma,\sigma}^{view}$ is defined implicitly in terms of the yet-to-be-determined view-implied expectations $\mathbb{E}^{f_{\boldsymbol{X}}}\{\boldsymbol{\zeta}_{\sigma}\boldsymbol{X}\} = \boldsymbol{\zeta}_{\sigma}\boldsymbol{\mu}$ as follows

$$\boldsymbol{\eta}_{\sigma,\sigma}^{view} \equiv \boldsymbol{\sigma}^{2view} + (\boldsymbol{\zeta}_{\sigma}\boldsymbol{\mu})(\boldsymbol{\zeta}_{\sigma}\boldsymbol{\mu})'. \tag{1.53}$$

For this reason, under views on covariances as in (1.51), the updated distribution (17) is still normal (1.15) and no recursion is needed. Indeed, the view-implied expectation (1.40) becomes explicit

$$\boldsymbol{\eta}_{\sigma}^{view} \equiv \mathbb{E}^{\bar{f}_{\boldsymbol{X}}}\{\boldsymbol{\zeta}_{\sigma}\boldsymbol{X}\} = \boldsymbol{\sigma}^{2view}(\boldsymbol{\zeta}_{\sigma}\underline{\boldsymbol{\sigma}}_{\boldsymbol{X}}^{2}\boldsymbol{\zeta}_{\sigma}')^{-1}\boldsymbol{\zeta}_{\sigma}\underline{\boldsymbol{\mu}}_{\boldsymbol{X}}, \tag{1.54}$$

and then both Lagrange multipliers $\boldsymbol{\theta}_{\sigma,\sigma}^{view}$ (1.36), updated expectation $\bar{\boldsymbol{\mu}}_{\boldsymbol{X}}$ (1.37) and covariance $\bar{\boldsymbol{\sigma}}_{\boldsymbol{X}}^{2}$ (1.38) follow in turn, as summarized in the table below.

Optimal Lagr. mult.	
$\boldsymbol{\theta}_{\mu}^{view} = \varnothing$	$\boldsymbol{\theta}_{\sigma,\sigma}^{view} = \frac{1}{2}((\boldsymbol{\zeta}_{\sigma}\underline{\boldsymbol{\sigma}}_{\boldsymbol{X}}^{2}\boldsymbol{\zeta}_{\sigma}')^{-1} - (\boldsymbol{\sigma}^{2view})^{-1})$
Updated exp. and cov.	
$\bar{\boldsymbol{\mu}}_{\boldsymbol{X}} = \underline{\boldsymbol{\mu}}_{\boldsymbol{X}} + \boldsymbol{\zeta}_{\sigma}^{\dagger\prime}(\boldsymbol{\sigma}^{2view}(\boldsymbol{\zeta}_{\sigma}\underline{\boldsymbol{\sigma}}_{\boldsymbol{X}}^{2}\boldsymbol{\zeta}_{\sigma}')^{-1} - \mathbb{I}_{\bar{k}\times\bar{k}})\boldsymbol{\zeta}_{\sigma}\underline{\boldsymbol{\mu}}_{\boldsymbol{X}}$	$\bar{\boldsymbol{\sigma}}_{\boldsymbol{X}}^{2} = \underline{\boldsymbol{\sigma}}_{\boldsymbol{X}}^{2} + \boldsymbol{\zeta}_{\sigma}^{\dagger\prime}(\boldsymbol{\sigma}^{2view} - \boldsymbol{\zeta}_{\sigma}\underline{\boldsymbol{\sigma}}_{\boldsymbol{X}}^{2}\boldsymbol{\zeta}_{\sigma}')\boldsymbol{\zeta}_{\sigma}^{\dagger}$

Table 1.4: Views on second moments: MRE solutions under views on covariances

Note that the updated expectation in Table 1.4 is consistent with its counterpart under equality views on first moments (1.21) and can be interpreted again as a projection similar to (1.26). Indeed, consider the $(\bar{n}-\bar{k})$-hyperplane $\mathcal{S}_{\boldsymbol{\zeta}_{\sigma}}$ defined as follows

$$\mathcal{S}_{\boldsymbol{\zeta}_{\sigma}} \equiv \{\boldsymbol{\mu} \in \mathbb{R}^{\bar{n}} \text{ such that } \boldsymbol{\zeta}_{\sigma}\boldsymbol{\mu} = \boldsymbol{0}\}, \tag{1.55}$$

and the projection operator of an $\bar{n}\times 1$ vector $\boldsymbol{x}$ onto $\mathcal{S}_{\boldsymbol{\zeta}_{\sigma}}$

$$\mathfrak{P}_{\sigma}[\boldsymbol{x}] = (\mathbb{I}_{\bar{n}} - \boldsymbol{\zeta}_{\sigma}^{\dagger\prime}\boldsymbol{\zeta}_{\sigma})\boldsymbol{x}, \tag{1.56}$$

which because of (1.39) is orthogonal with respect to the inner product $\langle \boldsymbol{x}, \boldsymbol{y}\rangle_{\boldsymbol{\omega}^2} \equiv \boldsymbol{x}'\boldsymbol{\omega}^2\boldsymbol{y}$ induced by either i) the inverse base covariance $\boldsymbol{\omega}^2 \equiv (\underline{\boldsymbol{\sigma}}_{\boldsymbol{X}}^{2})^{-1}$, or ii) the inverse updated covariance $\boldsymbol{\omega}^2 \equiv (\bar{\boldsymbol{\sigma}}_{\boldsymbol{X}}^{2})^{-1}$ (1.38).

Then, the updated expectation in Table 1.4 reads [A.1.7]

$$\bar{\boldsymbol{\mu}}_{\boldsymbol{X}} = \boldsymbol{\zeta}_{\sigma}^{\dagger\prime}\boldsymbol{\eta}_{\sigma}^{view} + \mathfrak{P}_{\sigma}[\underline{\boldsymbol{\mu}}_{\boldsymbol{X}} - \boldsymbol{\zeta}_{\sigma}^{\dagger\prime}\boldsymbol{\eta}_{\sigma}^{view}]; \tag{1.57}$$

and similar formulation follows for the updated covariance (1.38), up to vectorization

$$vec(\bar{\boldsymbol{\sigma}}_{\boldsymbol{X}}^{2}) = \boldsymbol{\zeta}_{\sigma,\sigma}^{\dagger\prime} vec(\boldsymbol{\sigma}^{2view}) + \mathfrak{P}_{\sigma,\sigma}[vec(\underline{\boldsymbol{\sigma}}_{\boldsymbol{X}}^{2}) - \boldsymbol{\zeta}_{\sigma,\sigma}^{\dagger\prime} vec(\boldsymbol{\sigma}^{2view})]. \tag{1.58}$$

Example 1.4. Consider the same setup as in Example 1.2. Then, if we input the true $\bar{k}\times 1$ vector of features (1.19) implied by the true updated distribution

$$\boldsymbol{\sigma}^{2view} \equiv \boldsymbol{\zeta}_{\sigma}\bar{\boldsymbol{\sigma}}^{2*}\boldsymbol{\zeta}_{\sigma}' = \left(\begin{smallmatrix} 0.19 & -0.17 \\ -0.17 & 0.29 \end{smallmatrix}\right), \tag{1.59}$$

in the formulation of the optimal Lagrange multipliers $\boldsymbol{\theta}_{\sigma,\sigma}^{view}$ in Table 1.4, we obtain back the true counterpart

$$||\boldsymbol{\theta}_{\sigma,\sigma}^{view*} - \boldsymbol{\theta}_{\sigma,\sigma}^{view}|| = 9.93 \times 10^{-16}. \tag{1.60}$$

Similar results follow for expectation $\bar{\boldsymbol{\mu}}_{\boldsymbol{X}}$ and covariance $\bar{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}$ in Table 1.4.

1.5 Views on first and second moments

Let us now suppose the general case of joint equality views (18) on linear combinations of expectations and second non-central moments as in (1.9), where $\boldsymbol{\zeta}_\mu$ and $\boldsymbol{\zeta}_\sigma$ are a $\bar{k}_\mu \times \bar{n}$ and a $\bar{k}_\sigma \times \bar{n}$ full rank matrices respectively. Then, similar to the case of views on second non-central moments (1.33), it turns out that the optimal Lagrange multipliers $\boldsymbol{\theta}_{\mu}^{view}$ (1.16) are defined implicitly in terms of the updated expectation of the view variables (1.34) [A.1.8]

$$\boldsymbol{\theta}_{\mu}^{view} = (\boldsymbol{\zeta}_\mu \bar{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} \boldsymbol{\zeta}'_\mu)^{-1}(\boldsymbol{\eta}_{\mu}^{view} - \boldsymbol{\zeta}_\mu \bar{\boldsymbol{\mu}}_{\boldsymbol{X};\sigma}), \tag{1.61}$$

where $\bar{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}$ is the updated covariance (1.38); $\bar{\boldsymbol{\mu}}_{\boldsymbol{X};\sigma}$ is the following $\bar{n} \times 1$ vector

$$\bar{\boldsymbol{\mu}}_{\boldsymbol{X};\sigma} \equiv \underline{\boldsymbol{\mu}}_{\boldsymbol{X}} + \boldsymbol{\zeta}_\sigma^{\dagger\prime}(\sigma^{2view}(\boldsymbol{\eta}_{\sigma}^{view})(\boldsymbol{\zeta}_\sigma \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} \boldsymbol{\zeta}'_\sigma)^{-1}\boldsymbol{\zeta}_\sigma \underline{\boldsymbol{\mu}}_{\boldsymbol{X}} - \boldsymbol{\zeta}_\sigma \underline{\boldsymbol{\mu}}_{\boldsymbol{X}}); \tag{1.62}$$

and $\sigma^{2view}(\boldsymbol{\eta}_{\sigma}^{view}) \equiv \boldsymbol{\eta}_{\sigma,\sigma}^{view} - \boldsymbol{\eta}_{\sigma}^{view}\boldsymbol{\eta}_{\sigma}^{view\prime}$ (1.35); and similar for the optimal Lagrange multipliers $\boldsymbol{\theta}_{\sigma,\sigma}^{view}$ (1.16), which reads as in (1.36)

$$\boldsymbol{\theta}_{\sigma,\sigma}^{view} = \frac{1}{2}((\boldsymbol{\zeta}_\sigma \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} \boldsymbol{\zeta}'_\sigma)^{-1} - (\sigma^{2view}(\boldsymbol{\eta}_{\sigma}^{view}))^{-1}). \tag{1.63}$$

Similar to the above, from the updated canonical coordinates (1.5), we can deduce also the updated expectation (1.17) [A.1.8]

$$\bar{\boldsymbol{\mu}}_{\boldsymbol{X}} = \bar{\boldsymbol{\mu}}_{\boldsymbol{X};\sigma} + \bar{\boldsymbol{\zeta}}_\mu^{\dagger\prime}(\boldsymbol{\eta}_{\mu}^{view} - \boldsymbol{\zeta}_\mu \bar{\boldsymbol{\mu}}_{\boldsymbol{X};\sigma}), \tag{1.64}$$

where $\bar{\boldsymbol{\zeta}}_\mu^{\dagger}$ is a $\bar{k}_\mu \times \bar{n}$ pseudo-inverse matrix for $\boldsymbol{\zeta}'_\mu$

$$\bar{\boldsymbol{\zeta}}_\mu^{\dagger} \equiv (\boldsymbol{\zeta}_\mu \bar{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} \boldsymbol{\zeta}'_\mu)^{-1}\boldsymbol{\zeta}_\mu \bar{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}; \tag{1.65}$$

and the updated covariance (1.38) reads as in (1.38)

$$\bar{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} = \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} + \boldsymbol{\zeta}_\sigma^{\dagger\prime}(\sigma^{2view}(\boldsymbol{\eta}_{\sigma}^{view}) - \boldsymbol{\zeta}_\sigma \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} \boldsymbol{\zeta}'_\sigma)\boldsymbol{\zeta}_\sigma^{\dagger}. \tag{1.66}$$

Note how $\bar{\boldsymbol{\mu}}_{\boldsymbol{X};\sigma}$ is consistent with the updated expectation under only views on second moments (1.37)-(1.40). Indeed, in the case of no views on expectations or $\boldsymbol{\zeta}_\mu = \mathbf{0}$, from (1.64) we would obtain $\bar{\boldsymbol{\mu}}_{\boldsymbol{X}} = \bar{\boldsymbol{\mu}}_{\boldsymbol{X};\sigma}$.

In conclusion, similar to the case of views on second moments in Section 1.4, when we consider views on first and second moments as in (18), *not* only the covariance, *but* also the expectation is updated.

1.5.1 Numerical solution via recursion

Similar to the case of views on second non-central moments (1.40), we can address the MRE solutions (1.61)-(1.63)-(1.64)-(1.66) numerically through a simple fixed-point recursion in the view-implied expectation (1.34) as follows from (1.64) [A.1.8]

$$\boldsymbol{\eta}_{\sigma}^{view} = g(\boldsymbol{\eta}_{\sigma}^{view}) \equiv (\boldsymbol{\eta}_{\sigma,\sigma}^{view} - \boldsymbol{\eta}_{\sigma}^{view}\boldsymbol{\eta}_{\sigma}^{view\prime})(\boldsymbol{\zeta}_{\sigma}\underline{\boldsymbol{\sigma}}_{\boldsymbol{X}}^{2}\boldsymbol{\zeta}_{\sigma}^{\prime})^{-1}\boldsymbol{\zeta}_{\sigma}\underline{\boldsymbol{\mu}}_{\boldsymbol{X}} + \boldsymbol{\zeta}_{\sigma}\bar{\boldsymbol{\zeta}}_{\mu}^{\dagger\prime}(\boldsymbol{\eta}_{\mu}^{view} - \boldsymbol{\zeta}_{\mu}\bar{\boldsymbol{\mu}}_{\boldsymbol{X};\sigma})), \quad (1.67)$$

where $\bar{\boldsymbol{\mu}}_{\boldsymbol{X};\sigma}$ is defined in (1.62). Note how (1.67) generalizes the recursion in the case of only views on second moments (1.40).

Then, the routine can be set up as follows.

$(\boldsymbol{\theta}_{\mu}^{view}, \boldsymbol{\theta}_{\sigma,\sigma}^{view}, \bar{\boldsymbol{\mu}}_{\boldsymbol{X}}, \bar{\boldsymbol{\sigma}}_{\boldsymbol{X}}^{2}) \leftarrow Fit.MRE.First.Second.Moments.N(\boldsymbol{\zeta}_{\mu}, \boldsymbol{\zeta}_{\sigma}, \boldsymbol{\eta}_{\mu}^{view}, \boldsymbol{\eta}_{\sigma,\sigma}^{view}, \underline{\boldsymbol{\mu}}_{\boldsymbol{X}}, \underline{\boldsymbol{\sigma}}_{\boldsymbol{X}}^{2})$	
0. Initialize	$\boldsymbol{\eta}_{\sigma}^{view} \leftarrow \boldsymbol{\zeta}_{\sigma}\underline{\boldsymbol{\mu}}_{\boldsymbol{X}}$
1. Update feat.	$\boldsymbol{\eta}_{\sigma}^{view} \leftarrow g(\boldsymbol{\eta}_{\sigma}^{view})$ (1.67)
2. If convergence, output $(\boldsymbol{\theta}_{\mu}^{view}, \boldsymbol{\theta}_{\sigma,\sigma}^{view}, \bar{\boldsymbol{\mu}}_{\boldsymbol{X}}, \bar{\boldsymbol{\sigma}}_{\boldsymbol{X}}^{2})$ (1.61)-(1.63)-(1.64)-(1.66); else go to 1	

Table 1.5: Iterative routine for optimal Lagrange multipliers under views on expectations and second non-central moments

Even here, If the true stationary point $\boldsymbol{\eta}_{\sigma}^{view}$ is an *attractive* fixed point for (1.40), the convergence in the above routine occurs when the relative norm between two subsequent updates $\boldsymbol{\eta}_{\sigma}^{view(i)}$ and $\boldsymbol{\eta}_{\sigma}^{view(i+1)}$ is smaller than a required threshold. See [Coxeter, 1998] for more details.

Example 1.5. Consider the same base model as in Example 1.1 and assume the true updated distribution $\boldsymbol{X} \sim Exp(\boldsymbol{\theta}_{\mu}^{view*}, \zeta^{view}, \underline{f}_{\boldsymbol{X}})$ (1.15) to be steered by the following Lagrange multipliers

$$\boldsymbol{\theta}_{\mu}^{view*} \equiv \begin{pmatrix} \theta_{\mu}^{view*} \\ \theta_{\sigma,\sigma}^{view*} \end{pmatrix} = \begin{pmatrix} 1.73 \\ 3.04 \end{pmatrix}; \quad (1.68)$$

with $\bar{k} = 2$ view variables specified by $\zeta^{view}(\boldsymbol{x})$ as in (1.3), where

$$\boldsymbol{\zeta}_{\mu} \equiv \begin{pmatrix} 1 & -1 & 0 \end{pmatrix}, \quad \boldsymbol{\zeta}_{\sigma} \equiv \begin{pmatrix} 0 & 1 & -1 \end{pmatrix}. \quad (1.69)$$

Then, if we input the true $\bar{k} \times 1$ vector of features (1.19) implied by the true updated distribution

$$\begin{aligned} \boldsymbol{\eta}^{view} &\equiv \begin{pmatrix} \eta_{\mu}^{view} \\ \eta_{\sigma,\sigma}^{view} \end{pmatrix} \\ &\equiv \begin{pmatrix} \zeta_{\mu}\bar{\mu}_{\boldsymbol{X}} \\ \zeta_{\sigma}(\bar{\boldsymbol{\sigma}}_{\boldsymbol{X}}^{2} + \bar{\mu}_{\boldsymbol{X}}\bar{\mu}_{\boldsymbol{X}}^{\prime})\zeta_{\sigma}^{\prime} \end{pmatrix} = \begin{pmatrix} 1.02 \\ 2.67 \end{pmatrix}, \end{aligned} \quad (1.70)$$

in the recursion in Table 1.5, we obtain back from the optimal Lagrange multipliers $\boldsymbol{\theta}^{view*} \equiv (\theta_{\mu}^{view*}, \theta_{\sigma,\sigma}^{view*})'$ (1.61)-(1.63) the true counterpart

$$||\boldsymbol{\theta}^{view*} - \boldsymbol{\theta}^{view}|| = 2.94 \times 10^{-7}. \tag{1.71}$$

Similar results follow for expectation $\bar{\boldsymbol{\mu}}_{\boldsymbol{X}}$ (1.64) and covariance $\bar{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}$ (1.66).

1.5.2 Analytical solution under special assumptions

We can address the MRE solutions (1.61)-(1.63)-(1.64)-(1.66) analytically under special assumptions on the base distribution and views.

Same view variables

Let us suppose the case of joint equality views (18) on *same* linear combinations ($\boldsymbol{\zeta}_{\mu} = \boldsymbol{\zeta}_{\sigma}$) of expectations and second non-central moments (1.9), or

$$f_{\boldsymbol{X}} \in \mathcal{C}_{\boldsymbol{X}} : \quad \begin{cases} \mathbb{E}^{f_{\boldsymbol{X}}}\{\boldsymbol{\zeta}\boldsymbol{X}\} = \boldsymbol{\eta}_{\mu}^{view} \\ \mathbb{E}^{f_{\boldsymbol{X}}}\{\boldsymbol{\zeta}\boldsymbol{X}\boldsymbol{X}'\boldsymbol{\zeta}\} = \boldsymbol{\eta}_{\sigma,\sigma}^{view} \end{cases} \tag{1.72}$$

where $\boldsymbol{\zeta}$ is a $\bar{k} \times \bar{n}$ full rank matrix.

Then no recursion is needed. Indeed, the view-implied expectation (1.67) is explicit by construction

$$\mathbb{E}^{\bar{f}_{\boldsymbol{X}}}\{\boldsymbol{\zeta}\boldsymbol{X}\} = \boldsymbol{\eta}_{\mu}^{view}, \tag{1.73}$$

and then both Lagrange multipliers $\boldsymbol{\theta}_{\mu}^{view}$ (1.61), $\boldsymbol{\theta}_{\sigma,\sigma}^{view}$ (1.63), updated expectation $\bar{\boldsymbol{\mu}}_{\boldsymbol{X}}$ (1.64) and covariance $\bar{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}$ (1.66) follow in turn, as summarized in the table below [A.1.9].

Optimal Lagr. mult.	
$\boldsymbol{\theta}_{\mu}^{view} = (\boldsymbol{\eta}_{\sigma,\sigma}^{view} - \boldsymbol{\eta}_{\mu}^{view}\boldsymbol{\eta}_{\mu}^{view\prime})^{-1}\boldsymbol{\eta}_{\mu}^{view} - (\boldsymbol{\zeta}\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}\boldsymbol{\zeta}')^{-1}\boldsymbol{\zeta}\underline{\boldsymbol{\mu}}_{\boldsymbol{X}}$	$\boldsymbol{\theta}_{\sigma,\sigma}^{view} = \frac{1}{2}((\boldsymbol{\zeta}\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}\boldsymbol{\zeta}')^{-1} - (\boldsymbol{\eta}_{\sigma,\sigma}^{view} - \boldsymbol{\eta}_{\mu}^{view}\boldsymbol{\eta}_{\mu}^{view\prime})^{-1})$
Updated exp. and cov.	
$\bar{\boldsymbol{\mu}}_{\boldsymbol{X}} = \underline{\boldsymbol{\mu}}_{\boldsymbol{X}} + \boldsymbol{\zeta}^{\dagger\prime}(\boldsymbol{\eta}_{\mu}^{view} - \boldsymbol{\zeta}\underline{\boldsymbol{\mu}}_{\boldsymbol{X}})$	$\bar{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} = \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} + \boldsymbol{\zeta}^{\dagger\prime}(\boldsymbol{\eta}_{\sigma,\sigma}^{view} - \boldsymbol{\eta}_{\mu}^{view}\boldsymbol{\eta}_{\mu}^{view\prime} - \boldsymbol{\zeta}\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}\boldsymbol{\zeta}')\boldsymbol{\zeta}^{\dagger}$

Table 1.6: Views on first two moments: MRE solutions under same combinations

This is consistent with [Qian and Gorman, 2001] which generalizes [Mina and Xiao, 2001]. Indeed the views (1.72) are equivalent to equality views (18) on linear combinations of expectations and covariances

$$f_{\boldsymbol{X}} \in \mathcal{C}_{\boldsymbol{X}} : \quad \begin{cases} \mathbb{E}^{f_{\boldsymbol{X}}}\{\boldsymbol{\zeta}\boldsymbol{X}\} = \boldsymbol{\mu}^{view} \\ \mathbb{C}v^{f_{\boldsymbol{X}}}\{\boldsymbol{\zeta}\boldsymbol{X}\} = \boldsymbol{\sigma}^{2view} \end{cases} \tag{1.74}$$

up to redefine the view features as follows

$$\boldsymbol{\eta}_{\mu}^{view} \equiv \boldsymbol{\mu}^{view}, \quad \boldsymbol{\eta}_{\sigma,\sigma}^{view} \equiv \boldsymbol{\sigma}^{2view} + \boldsymbol{\eta}_{\mu}^{view}\boldsymbol{\eta}_{\mu}^{view\prime}. \tag{1.75}$$

Example 1.6. Consider the same base model as in Example 1.1 and assume the true updated distribution $\boldsymbol{X} \sim Exp(\boldsymbol{\theta}^{view*}, \zeta^{view}, \underline{f}_{\boldsymbol{X}})$ (1.15) to be steered by the following Lagrange multipliers

$$\boldsymbol{\theta}_{\mu}^{view*} = \begin{pmatrix} 5.71 \\ 0.38 \end{pmatrix}, \quad \boldsymbol{\theta}_{\sigma,\sigma}^{view*} = \begin{pmatrix} -1.83 & -2.82 \\ -2.82 & -3.13 \end{pmatrix}; \tag{1.76}$$

with $\bar{k} = 6$ view variables specified by $\zeta^{view}(\boldsymbol{x})$ as in (1.3) on the same linear combination

$$\boldsymbol{\zeta} \equiv \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix}. \tag{1.77}$$

Then, if we input the true $\bar{k} \times 1$ vector of features (1.19) implied by the true updated distribution

$$\boldsymbol{\eta}_{\mu}^{view} = \boldsymbol{\zeta}\bar{\boldsymbol{\mu}}_{\boldsymbol{X}} = \begin{pmatrix} 1.18 \\ -0.76 \end{pmatrix}, \quad \boldsymbol{\eta}_{\sigma,\sigma}^{view} = \boldsymbol{\zeta}(\bar{\boldsymbol{\sigma}}_{\boldsymbol{X}}^{2} + \bar{\boldsymbol{\mu}}_{\boldsymbol{X}}\bar{\boldsymbol{\mu}}_{\boldsymbol{X}}')\boldsymbol{\zeta}' = \begin{pmatrix} 1.62 & -1.04 \\ -1.04 & 0.73 \end{pmatrix}, \tag{1.78}$$

in the formulations of the optimal Lagrange multipliers $\boldsymbol{\theta}^{view} \equiv (\boldsymbol{\theta}_{\mu}^{view}, vec(\boldsymbol{\theta}_{\sigma,\sigma}^{view}))'$ in Table 1.6, we obtain back the true counterpart

$$||\boldsymbol{\theta}^{view*} - \boldsymbol{\theta}^{view}|| = 6.15 \times 10^{-15}. \tag{1.79}$$

Similar results follow for expectation $\bar{\boldsymbol{\mu}}_{\boldsymbol{X}}$ and covariance $\bar{\boldsymbol{\sigma}}_{\boldsymbol{X}}^{2}$ in Table 1.6.

Views on expectations and covariances

Let us suppose the case of joint equality views (18) on linear combinations of expectations and covariances

$$f_{\boldsymbol{X}} \in \mathcal{C}_{\boldsymbol{X}}: \quad \begin{cases} \mathbb{E}^{f_{\boldsymbol{X}}}\{\boldsymbol{\zeta}_{\mu}\boldsymbol{X}\} = \boldsymbol{\mu}^{view} \\ \mathbb{C}v^{f_{\boldsymbol{X}}}\{\boldsymbol{\zeta}_{\sigma}\boldsymbol{X}\} = \boldsymbol{\sigma}^{2view} \end{cases} \tag{1.80}$$

Then, similar to the case of views of covariances (1.51), the views can be easily re-written as linear combinations of expectations and second non-central moments (1.9)

$$f_{\boldsymbol{X}} \in \mathcal{C}_{\boldsymbol{X}}: \quad \begin{cases} \mathbb{E}^{f_{\boldsymbol{X}}}\{\boldsymbol{\zeta}_{\mu}\boldsymbol{X}\} = \boldsymbol{\eta}_{\mu}^{view} \\ \mathbb{E}^{f_{\boldsymbol{X}}}\{\boldsymbol{\zeta}_{\sigma}\boldsymbol{X}\boldsymbol{X}'\boldsymbol{\zeta}_{\sigma}\} = \boldsymbol{\eta}_{\sigma,\sigma}^{view} \end{cases} \tag{1.81}$$

where the $\bar{k}_{\sigma} \times \bar{k}_{\sigma}$ matrix of features $\boldsymbol{\eta}_{\sigma,\sigma}^{view}$ is defined implicitly in terms of the yet-to-be-determined view-implied expectations $\mathbb{E}^{f_{\boldsymbol{X}}}\{\boldsymbol{\zeta}_{\sigma}\boldsymbol{X}\} = \boldsymbol{\zeta}_{\sigma}\boldsymbol{\mu}$ as follows

$$\boldsymbol{\eta}_{\mu}^{view} \equiv \boldsymbol{\mu}^{view}, \quad \boldsymbol{\eta}_{\sigma,\sigma}^{view} \equiv \boldsymbol{\sigma}^{2view} + (\boldsymbol{\zeta}_{\sigma}\boldsymbol{\mu})(\boldsymbol{\zeta}_{\sigma}\boldsymbol{\mu})'. \tag{1.82}$$

For this reason, under views on expectations and covariances as in (1.80), the updated distribution (17) is still normal (1.15) and in these circumstances no recursion is needed. Indeed, the view-implied expectation (1.67) becomes explicit

$$\boldsymbol{\eta}_{\sigma}^{view} = \boldsymbol{\sigma}^{2view}(\boldsymbol{\zeta}_{\sigma}\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}\boldsymbol{\zeta}_{\sigma}')^{-1}\boldsymbol{\zeta}_{\sigma}\underline{\boldsymbol{\mu}}_{\boldsymbol{X}} + \boldsymbol{\zeta}_{\sigma}\bar{\boldsymbol{\zeta}}_{\mu}^{\dagger\prime}(\boldsymbol{\eta}_{\mu}^{view} - \boldsymbol{\zeta}_{\mu}\bar{\boldsymbol{\mu}}_{\boldsymbol{X};\sigma}), \tag{1.83}$$

where $\bar{\boldsymbol{\mu}}_{\boldsymbol{X};\sigma}$ (1.62) becomes

$$\bar{\boldsymbol{\mu}}_{\boldsymbol{X};\sigma} = \underline{\boldsymbol{\mu}}_{\boldsymbol{X}} + \boldsymbol{\zeta}_{\sigma}^{\dagger\prime}(\boldsymbol{\sigma}^{2view}(\boldsymbol{\zeta}_{\sigma}\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}\boldsymbol{\zeta}_{\sigma}')^{-1}\boldsymbol{\zeta}_{\sigma}\underline{\boldsymbol{\mu}}_{\boldsymbol{X}} - \boldsymbol{\zeta}_{\sigma}\underline{\boldsymbol{\mu}}_{\boldsymbol{X}}), \tag{1.84}$$

and then both Lagrange multipliers $\boldsymbol{\theta}_{\mu}^{view}$ (1.61), $\boldsymbol{\theta}_{\sigma,\sigma}^{view}$ (1.63), updated expectation $\bar{\boldsymbol{\mu}}_{\boldsymbol{X}}$ (1.64) and covariance $\bar{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}$ (1.66) follow in turn, as summarized in the table below [A.1.10].

Optimal Lagr. mult.	
$\boldsymbol{\theta}_{\mu}^{view} = (\boldsymbol{\zeta}_{\mu}\bar{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}\boldsymbol{\zeta}_{\mu}')^{-1}(\boldsymbol{\mu}^{view} - \boldsymbol{\zeta}_{\mu}\bar{\boldsymbol{\mu}}_{\boldsymbol{X};\sigma})$	$\boldsymbol{\theta}_{\sigma,\sigma}^{view} = \frac{1}{2}((\boldsymbol{\zeta}_{\sigma}\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}\boldsymbol{\zeta}_{\sigma}')^{-1} - (\boldsymbol{\sigma}^{2view})^{-1})$
Updated exp. and cov.	
$\bar{\boldsymbol{\mu}}_{\boldsymbol{X}} = \bar{\boldsymbol{\mu}}_{\boldsymbol{X};\sigma} + \bar{\boldsymbol{\zeta}}_{\mu}^{\dagger\prime}(\boldsymbol{\mu}^{view} - \boldsymbol{\zeta}_{\mu}\bar{\boldsymbol{\mu}}_{\boldsymbol{X};\sigma})$	$\bar{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} = \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} + \boldsymbol{\zeta}_{\sigma}^{\dagger\prime}(\boldsymbol{\sigma}^{2view} - \boldsymbol{\zeta}_{\sigma}\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}\boldsymbol{\zeta}_{\sigma}')\boldsymbol{\zeta}_{\sigma}^{\dagger}$

Table 1.7: Views on first two moments: MRE solutions under views on expectations and covariances

In particular, the updated expectation is again consistent with its counterpart under equality views on first moments (1.21) and can be interpreted as a sequential projection: first onto the $(\bar{n} - \bar{k}_{\sigma})$-hyperplane $\mathcal{S}_{\boldsymbol{\zeta}_{\sigma}}$ (1.56) and then onto the $(\bar{n} - \bar{k}_{\mu})$-hyperplane $\mathcal{S}_{\boldsymbol{\zeta}_{\mu}}$ (1.24)

$$\bar{\boldsymbol{\mu}}_{\boldsymbol{X}} = \bar{\boldsymbol{\zeta}}_{\mu}^{\dagger\prime}\boldsymbol{\eta}_{\mu}^{view} + \mathfrak{P}_{\mu}[\boldsymbol{\zeta}_{\sigma}^{\dagger\prime}\boldsymbol{\eta}_{\sigma}^{view} + \mathfrak{P}_{\sigma}[\underline{\boldsymbol{\mu}}_{\boldsymbol{X}} - \boldsymbol{\zeta}_{\sigma}^{\dagger\prime}\boldsymbol{\eta}_{\sigma}^{view}] - \bar{\boldsymbol{\zeta}}_{\mu}^{\dagger\prime}\boldsymbol{\eta}_{\mu}^{view}]. \tag{1.85}$$

Note how in this case the covariances $\bar{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}$ and $\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}$ have a significant role in both projections. This because the sequential projection is *not* commutative, as showed in Figure 1.2 and highlighted also in the definition of pseudo inverse $\bar{\boldsymbol{\zeta}}_{\mu}^{\dagger}$ (1.65), which is weighted according to the *updated* covariance $\bar{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}$.

As a matter of fact, differently from the case of only views on first moments, the pseudo inverses $\bar{\boldsymbol{\zeta}}_{\mu}^{\dagger}$ and $\boldsymbol{\zeta}_{\mu}^{\dagger}$ (1.22) are *different*

$$\bar{\boldsymbol{\zeta}}_{\mu}^{\dagger} \neq \boldsymbol{\zeta}_{\mu}^{\dagger}. \tag{1.86}$$

Example 1.7. Consider the same case study as in Example 1.5. Then, if we input the features (1.80) implied by the true updated distribution

$$\mu^{view} \equiv \eta_{\mu}^{view} = 1.02, \quad \sigma^{2view} \equiv \boldsymbol{\zeta}_{\sigma}\bar{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}\boldsymbol{\zeta}_{\sigma}' = 0.94, \tag{1.87}$$

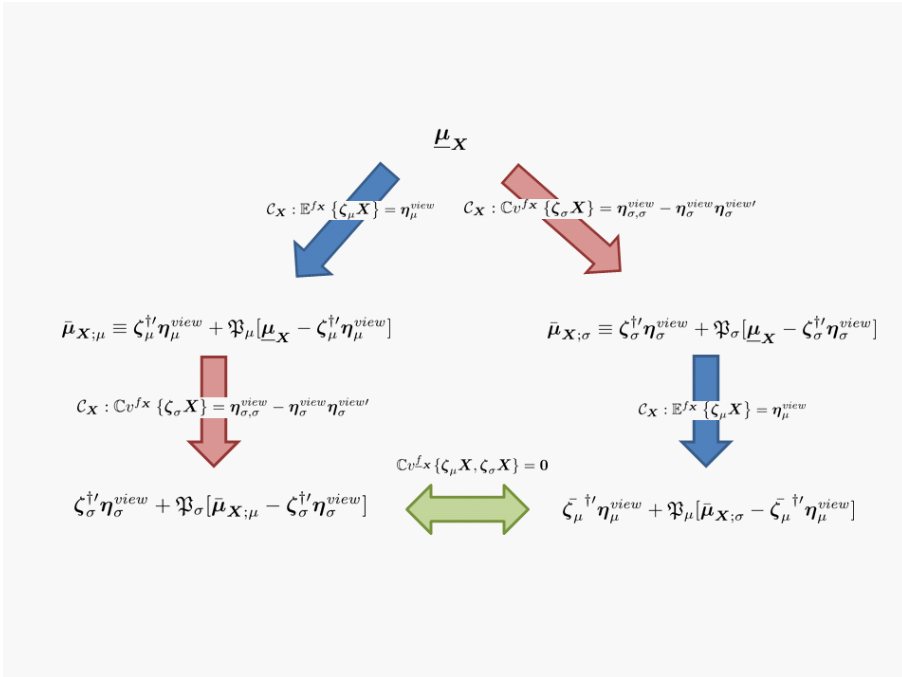

Figure 1.2: Views on expectations and covariances: updated expectation via sequential projections

in the formulations of the optimal Lagrange multipliers $\boldsymbol{\theta}^{view} \equiv (\theta_\mu^{view}, \theta_{\sigma,\sigma}^{view})'$ in Table 1.7, we obtain back the true counterparts

$$||\boldsymbol{\theta}^{view*} - \boldsymbol{\theta}^{view}|| = 9.93 \times 10^{-16}. \tag{1.88}$$

Similar results follow for expectation $\bar{\boldsymbol{\mu}}_{\boldsymbol{X}}$ and covariance $\bar{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}$ in Table 1.7.

However, *if* the view variables are statistically independent under the base distribution (1.6), or

$$\mathbb{C}v^{\underline{f}_{\boldsymbol{X}}}\{\boldsymbol{\zeta}_\mu \boldsymbol{X}, \boldsymbol{\zeta}_\sigma \boldsymbol{X}\} = \boldsymbol{\zeta}_\mu \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} \boldsymbol{\zeta}_\sigma' = \boldsymbol{0}, \tag{1.89}$$

then several properties follows [A.1.11]:
- the pseudo inverses $\bar{\boldsymbol{\zeta}}_\mu^{\dagger}$ and $\boldsymbol{\zeta}_\mu^{\dagger}$ becomes the same

$$\bar{\boldsymbol{\zeta}}_\mu^{\dagger} = \boldsymbol{\zeta}_\mu^{\dagger}; \tag{1.90}$$

- the projectors commutes

$$\mathfrak{P}_\mu[\mathfrak{P}_\sigma[\boldsymbol{x}]] = \mathfrak{P}_\sigma[\mathfrak{P}_\mu[\boldsymbol{x}]]; \tag{1.91}$$

- the view variables are statistically independent also under the updated distribution (1.15)

$$\mathbb{C}v^{\bar{f}_{\boldsymbol{X}}}\{\boldsymbol{\zeta}_\mu \boldsymbol{X}, \boldsymbol{\zeta}_\sigma \boldsymbol{X}\} = \boldsymbol{\zeta}_\mu \bar{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} \boldsymbol{\zeta}_\sigma' = \boldsymbol{0}; \tag{1.92}$$

- the updated parameters in Table 1.7 simplifies to the below [A.1.10]

Optimal Lagr. mult.	
$\boldsymbol{\theta}_{\mu}^{view} = (\boldsymbol{\zeta}_{\mu}\underline{\boldsymbol{\sigma}}_{\boldsymbol{X}}^{2}\boldsymbol{\zeta}_{\mu}')^{-1}(\boldsymbol{\mu}^{view} - \boldsymbol{\zeta}_{\mu}\underline{\boldsymbol{\mu}}_{\boldsymbol{X}})$	$\boldsymbol{\theta}_{\sigma,\sigma}^{view} = \frac{1}{2}((\boldsymbol{\zeta}_{\sigma}\underline{\boldsymbol{\sigma}}_{\boldsymbol{X}}^{2}\boldsymbol{\zeta}_{\sigma}')^{-1} - (\boldsymbol{\sigma}^{2view})^{-1})$
Updated exp. and cov.	
$\bar{\boldsymbol{\mu}}_{\boldsymbol{X}} = \bar{\boldsymbol{\mu}}_{\boldsymbol{X};\sigma} + \boldsymbol{\zeta}_{\mu}^{\dagger\prime}(\boldsymbol{\mu}^{view} - \boldsymbol{\zeta}_{\mu}\underline{\boldsymbol{\mu}}_{\boldsymbol{X}})$	$\bar{\boldsymbol{\sigma}}_{\boldsymbol{X}}^{2} = \underline{\boldsymbol{\sigma}}_{\boldsymbol{X}}^{2} + \boldsymbol{\zeta}_{\sigma}^{\dagger\prime}(\boldsymbol{\sigma}^{2view} - \boldsymbol{\zeta}_{\sigma}\underline{\boldsymbol{\sigma}}_{\boldsymbol{X}}^{2}\boldsymbol{\zeta}_{\sigma}')\boldsymbol{\zeta}_{\sigma}^{\dagger}$

Table 1.8: Views on first two moments: MRE solutions under views on expectations and covariances and independent views

This is consistent with [Meucci, 2010], up to a correction term in the updated expectation

$$\bar{\boldsymbol{\mu}}_{\boldsymbol{X}} = \underbrace{\underline{\boldsymbol{\mu}}_{\boldsymbol{X}} + \boldsymbol{\zeta}_{\mu}^{\dagger\prime}(\boldsymbol{\mu}^{view} - \boldsymbol{\zeta}_{\mu}\underline{\boldsymbol{\mu}}_{\boldsymbol{X}})}_{\text{mean-view update}} + \underbrace{\boldsymbol{\zeta}_{\sigma}^{\dagger\prime}(\boldsymbol{\sigma}^{2view}(\boldsymbol{\zeta}_{\sigma}\underline{\boldsymbol{\sigma}}_{\boldsymbol{X}}^{2}\boldsymbol{\zeta}_{\sigma}')^{-1}\boldsymbol{\zeta}_{\sigma}\underline{\boldsymbol{\mu}}_{\boldsymbol{X}} - \boldsymbol{\zeta}_{\sigma}\underline{\boldsymbol{\mu}}_{\boldsymbol{X}})}_{\text{mean correction}}. \tag{1.93}$$

Example 1.8. Consider the same case study as in Example 1.5, but with $\bar{k} = 2$ view variables specified by $\zeta^{view}(\boldsymbol{x})$ as in (1.3), where $\boldsymbol{\zeta}_{\mu}$ and $\boldsymbol{\zeta}_{\sigma}$ are specified as the first two eigenvectors of the base covariance $\underline{\boldsymbol{\sigma}}_{\boldsymbol{X}}^{2}$

$$\boldsymbol{\zeta}_{\mu} \equiv \left(0.43 \;\; 0.46 \;\; -0.78 \right), \quad \boldsymbol{\zeta}_{\sigma} \equiv \left(0.76 \;\; -0.65 \;\; 0.03 \right). \tag{1.94}$$

Then the view variables are statistically independent (1.89), since we have

$$\boldsymbol{\zeta}_{\mu}\underline{\boldsymbol{\sigma}}_{\boldsymbol{X}}^{2}\boldsymbol{\zeta}_{\sigma}' = 0. \tag{1.95}$$

In particular, if we input the features (1.80) implied by the true updated distribution

$$\mu^{view} \equiv \eta_{\mu}^{view} = 0.09, \quad \sigma^{2view} \equiv \boldsymbol{\zeta}_{\sigma}\bar{\boldsymbol{\sigma}}_{\boldsymbol{X}}^{2}\boldsymbol{\zeta}_{\sigma}' = 0.21, \tag{1.96}$$

in the formulations of the optimal Lagrange multipliers $\boldsymbol{\theta}^{view} \equiv (\theta_{\mu}^{view}, \theta_{\sigma,\sigma}^{view})'$ in Table 1.8, we obtain back the true counterparts

$$||\boldsymbol{\theta}^{view*} - \boldsymbol{\theta}^{view}|| = 2.23 \times 10^{-15}. \tag{1.97}$$

Similar results follow for expectation $\bar{\boldsymbol{\mu}}_{\boldsymbol{X}}$ and covariance $\bar{\boldsymbol{\sigma}}_{\boldsymbol{X}}^{2}$ in Table 1.8.

1.6 Conclusions

In this chapter we showed how to solve analytically the MRE problem under normality assumption of the base distribution and views on linear combinations of the first two moments.

In such a setting, we derived the explicit solutions of the updated expectation and covariance rephrasing the notable results of [Mina and Xiao, 2001], [Black and Litterman, 1990], [Qian and Gorman, 2001], [Meucci, 2010] in the more canonical parametrization via optimal Lagrange multipliers within the exponential-family class.

The main theoretical result is about conjugate distributions which shows, consistently with Bayesian theory [Murphy, 2007], how the exponential family distributions (and hence also the normal, as special case) are invariant under the MRE principle: when the base belongs to an exponential family class and the functions specifying the views are *linear* in the sufficient statistics, the MRE updated distribution belongs to the *same* exponential family class. In particular, under normality, we provided the formulations of the updated expectations and covariances, which are obtained either *numerically* via suitable fixed-point recursions, under views on *non-central moments*; or *analytically*, under views on *central moments*, much like in [Meucci, 2010].

Under normality, another relevant insight is the interpretation of MRE solution in terms of orthogonal projections which commute when the view variables are *statistically orthogonal*. In these circumstances we surprisingly found that the original updating rule for expectation derived in [Meucci, 2010] is imprecise and must be adjusted with an additional component implied by the views on covariances.

Chapter 2

Advanced portfolio construction via MRE approach

2.1 Introduction

In portfolio management, the traditional implementation a-la "Modern Portfolio Theory" pioneered by [Markowitz, 1952] is typically performed in two steps: i) estimate the moments of the market variables, such as returns, from realized historical data; and ii) plug-in the estimates, such as sample mean and covariance, as they were the true parameters in the mean-variance problem.

However, the estimation risk is central in this procedure. As matter of fact, [DeMiguel et al., 2009] illustrated how implementing optimal portfolios with sample moments gives rise to extreme fluctuating holdings, which perform poorly out of sample. For this reason a significant effort in finance has been devoted for handling the estimation error with the aim of improving the performance of the original portfolio selection model pioneered by [Markowitz, 1952]. In this context, modern literature shows how to reduce the impact of estimation error and the extent of possible outperformance over classical sample estimators via *Bayesian* or more general *Shrinkage* estimators, see for instance [Ben-Tal and Nemirovski, 2001], [Jagannathan and Ma, 2003], [Kan and Zhou, 2006], [Meucci, 2005].

Similar to the James-Stein estimator [Stein, 1955], the MRE framework provides a sound rationale for shrinkage estimates, which blend a pure statistical estimate, such as the historical variance, in order to satisfy additional constraints implied by the market views, such as target bounds on volatilities. MRE approach have already been extensively used in finance for applications including derivatives pricing ([Avellaneda, 1999], [D'Amico et al., 2003]), portfolio allocation ([Pezier, 2007]), stress-testing ([Breuer and Csiszar, 2013]), and, more broadly, in risk and portfolio management [Meucci, 2008], [Meucci, 2013], [Meucci, 2011], [Meucci, 2012b], [Meucci and Nicolosi, 2016].

A relevant application of the MRE is for the construction of a quantitative trading strategy, by processing ranking signals for alpha-generation. In the standard approach, discussed for instance in [Grinold and Kahn, 1999], the expected return of all the financial instruments in a given market is set proportional to the z-score of a given predictive signal. However, this procedure imposes restrictions in the optimization process. [Almgren and Chriss, 2006] address this issue, obtaining expected returns that do not overlay spurious informations. Instead, [Meucci et al., 2011] propose a parametric implementation of the MRE under a low-rank-diagonal structure of the covariance matrix, which is suitable for handling large dimensional markets and starts from the empirical observations. However this implementation presents several problems: the updated distribution is not represented in its (low-dimensional) canonical form via Lagrange multipliers, but parametrized directly in terms of (large-dimensional) expectation and covariance; the low-rank-diagonal parametrization of the covariance is affected by problems of identifications; the updated expectation and covariance are computed according to a non-convex optimization problem which does not admit a unique solution.

Here, we generalize the results of the parametric implementation of MRE in Chapter 1 and propose an alternative approach to [Meucci et al., 2011] for processing complex (in)equality views on first and second moments of the market variables, including views on ranking as special case.

The remainder of this chapter is organized as follows.

In Section 2.2 we review MRE theoretical framework for normal base distributions and (in)equality views on first two moments, which we then optimize numerically. In Section 2.3 we rephrase the MRE problem for (in)equality linear views on first moments as a linearly constrained quadratic programming problem. In Section 2.4 we review the more general case of (in)equality linear views on the first two second moments and derive the analytical expressions for the derivatives to feed in the optimization solver. In Section 2.5, we evaluate the performance, in terms of P&L Sharpe ratios, of optimal portfolio policies calibrated to U.S. stock market data in the Dow Jones Index, $\bar{n} = 30$ assets, which embed in the estimation process the views on ranking via: i) common approach by [Grinold and Kahn, 1999]; ii) Factor Entropy Pooling by [Meucci et al., 2011]; and iii) MRE, and compare the results with respect to the benchmark strategy of investing a fraction $1/\bar{n}$ of budget in each of the $\bar{n}$ financial instruments available. Finally, in Section 2.6 we list the main contributions.

Fully documented code is available on GitHub.

2.2 The model

Following the theoretical framework (4), here we address the MRE problem (16) under normality (1.6) and general (in)equality views. See also Figure 1.

More precisely, let us consider the following setup.

Suppose:
- a normal base distribution (1) as in the analytical approach (1.6)

$$\boldsymbol{X} \sim N(\underline{\boldsymbol{\mu}}_{\boldsymbol{X}}, \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}); \tag{2.1}$$

- (in)equality views on the first two moments of the market variables

$$f_{\boldsymbol{X}} \in \mathcal{C}_{\boldsymbol{X}}: \quad \begin{cases} \mathbb{E}^{f_{\boldsymbol{X}}}\{\boldsymbol{\zeta}_\mu \boldsymbol{X}\} \leqq \boldsymbol{\eta}_\mu^{view} \\ \mathbb{E}^{f_{\boldsymbol{X}}}\{\boldsymbol{\zeta}_\sigma \boldsymbol{X}\boldsymbol{X}'\boldsymbol{\zeta}_\sigma'\} \leqq \boldsymbol{\eta}_{\sigma,\sigma}^{view}. \end{cases} \tag{2.2}$$

Then, under normal base assumption (2.1) and inequality views (2.2), the MRE updated distribution (23) must be normal in turn as in the analytical counterpart (1.15) [A.1.3]

$$\boldsymbol{X} \sim N(\bar{\boldsymbol{\mu}}_{\boldsymbol{X}}, \bar{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}), \tag{2.3}$$

where the updated expectation (1.17) reads

$$\bar{\boldsymbol{\mu}}_{\boldsymbol{X}} = \mu(\boldsymbol{\theta}^{view}) \equiv \bar{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}((\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}})^{-1}\underline{\boldsymbol{\mu}}_{\boldsymbol{X}} + \boldsymbol{\zeta}_\mu'\boldsymbol{\theta}_\mu^{view}), \tag{2.4}$$

and the updated covariance (1.18) reads

$$\bar{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} = \sigma^2(\boldsymbol{\theta}^{view}) \equiv \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} + \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}\boldsymbol{\zeta}_\sigma'(\frac{1}{2}(\boldsymbol{\theta}_{\sigma,\sigma}^{view})^{-1} - \boldsymbol{\zeta}_\sigma\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}\boldsymbol{\zeta}_\sigma')^{-1}\boldsymbol{\zeta}_\sigma\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}. \tag{2.5}$$

In practice, $\boldsymbol{\theta}^{view}$ in (2.4)-(2.5) are the optimal Lagrange multipliers as in (27), which we report here

$$\boldsymbol{\theta}^{view} \equiv \begin{pmatrix} \boldsymbol{\theta}_\mu^{view} \\ vec(\boldsymbol{\theta}_{\sigma,\sigma}^{view}) \end{pmatrix} \equiv \underset{\boldsymbol{t}\leqq\boldsymbol{0}}{\operatorname{argmin}}\, \mathcal{L}(\boldsymbol{t};\boldsymbol{\eta}^{view}), \tag{2.6}$$

where dual Lagrangian (28) explicitly reads [A.1.3]

$$\mathcal{L}(\boldsymbol{t};\boldsymbol{\eta}^{view}) \equiv \psi^N(\mu(\boldsymbol{t}),\sigma^2(\boldsymbol{t})) - \psi^N(\underline{\boldsymbol{\mu}}_{\boldsymbol{X}},\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}) - \boldsymbol{t}_\mu'\boldsymbol{\eta}_\mu^{view} - \operatorname{tr}(\boldsymbol{t}_{\sigma,\sigma}'\boldsymbol{\eta}_{\sigma,\sigma}^{view}), \tag{2.7}$$

$\mu(\boldsymbol{t})$ is the $(\bar{n}\times 1)$-valued function defined as in (2.4); $\sigma^2(\boldsymbol{t})$ is the $(\bar{n}\times\bar{n})$-valued function defined as in (2.5); and where $\psi^N(\boldsymbol{\mu}_{\boldsymbol{X}},\boldsymbol{\sigma}^2_{\boldsymbol{X}})$ denotes the log-partition function of a multivariate normal distribution parametrized in terms of expectation and covariance [A.1.2]

$$\psi^N(\boldsymbol{\mu}_{\boldsymbol{X}},\boldsymbol{\sigma}^2_{\boldsymbol{X}}) \equiv \frac{1}{2}[\boldsymbol{\mu}_{\boldsymbol{X}}'(\boldsymbol{\sigma}^2_{\boldsymbol{X}})^{-1}\boldsymbol{\mu}_{\boldsymbol{X}} - \frac{1}{2}\ln\det((\boldsymbol{\sigma}^2_{\boldsymbol{X}})^{-1})]. \tag{2.8}$$

In these circumstances the dual Lagrangian problem (2.6) cannot be solved analytically as in Chapter 1.

However, as we proceed to discuss, since the large-dimensional location and dispersion $(\bar{\boldsymbol{\mu}}_{\boldsymbol{X}},\bar{\boldsymbol{\sigma}}^2_{\boldsymbol{X}})$ (2.4)-(2.5), which identify the updated distribution (2.3), are fully determined by a relatively *small* number $\bar{k}_\mu + \bar{k}_\sigma^2$ of parameters $\boldsymbol{\theta}^{view}$ (1.16), we can provide very efficiently the optimal solution (2.6) via numerical approximation.

2.3 (In)equality views on first moments

Let us suppose the case of only (in)equality views (18) on linear combinations of the first moments

$$f_{\boldsymbol{X}} \in \mathcal{C}_{\boldsymbol{X}} : \qquad \mathbb{E}^{f_{\boldsymbol{X}}}\{\boldsymbol{\zeta}_{\mu}\boldsymbol{X}\} \leqq \boldsymbol{\eta}_{\mu}^{view}, \tag{2.9}$$

where $\boldsymbol{\zeta}_{\mu}$ is a $\bar{k} \times \bar{n}$ full rank matrix.

Then the updated expectation (2.4) becomes

$$\bar{\boldsymbol{\mu}}_{\boldsymbol{X}} = \underline{\boldsymbol{\mu}}_{\boldsymbol{X}} + \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}\zeta'_{\mu}\boldsymbol{\theta}_{\mu}^{view}, \tag{2.10}$$

and the updated covariance (2.5) becomes as the base counterpart

$$\bar{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} = \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}. \tag{2.11}$$

Moreover, the dual Lagrangian (2.7) explicitly reads [A.2.2]

$$\mathcal{L}(\boldsymbol{t}_{\mu}; \boldsymbol{\eta}_{\mu}^{view}) = \frac{1}{2}\boldsymbol{t}'_{\mu}(\boldsymbol{\zeta}_{\mu}\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}\boldsymbol{\zeta}'_{\mu})\boldsymbol{t}_{\mu} + \boldsymbol{t}'_{\mu}(\boldsymbol{\zeta}_{\mu}\underline{\boldsymbol{\mu}}_{\boldsymbol{X}} - \boldsymbol{\eta}_{\mu}^{view}). \tag{2.12}$$

Hence the dual Lagrangian problem (2.6) is an instance of a linearly constrained *quadratic* programming problem, and as such, we can address it numerically via a built-in quadratic programming solver.

In particular, we can interpret the dual Lagrangian optimization (2.6) as in instance of mean-variance allocation problem.

Indeed, let us define the factor portfolios, or factors, as the excess view variables over the features

$$\boldsymbol{Z} \equiv \boldsymbol{\zeta}_{\mu}\boldsymbol{X} - \boldsymbol{\eta}_{\mu}^{view}. \tag{2.13}$$

Then we can re-write (2.6) as follows [A.2.3]

$$\boldsymbol{h}_{\lambda^*} \equiv \underset{\boldsymbol{h} \geqq \boldsymbol{0}}{\operatorname{argmax}}\, \underline{\mathbb{E}}\{Z_{\boldsymbol{h}}\} - \lambda^* \underline{\mathbb{V}}\{Z_{\boldsymbol{h}}\}, \tag{2.14}$$

where we defined $Z_{\boldsymbol{h}} \equiv \boldsymbol{h}'\boldsymbol{Z}$ and $\lambda^* \equiv \frac{1}{2}$; and where $\underline{\mathbb{E}}\{\cdot\}$ and $\underline{\mathbb{V}}\{\cdot\}$ denotes the expectation and variance operators respectively under the base distribution (2.1). The the optimal Lagrange multipliers $\boldsymbol{\theta}_{\mu}^{view}$ satisfies

$$\boldsymbol{\theta}_{\mu}^{view} = -\boldsymbol{h}_{\lambda^*}. \tag{2.15}$$

Example 2.1. Consider $\bar{n} \equiv 3$ market variables $\boldsymbol{X} \equiv (X_1, X_2, X_3)'$ with joint normal base distribution

$$\boldsymbol{X} \sim N(\underline{\boldsymbol{\mu}}_{\boldsymbol{X}}, \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}), \tag{2.16}$$

where

$$\underline{\boldsymbol{\mu}}_{\boldsymbol{X}} \equiv \begin{pmatrix} 0.26 \\ 0.29 \\ 0.33 \end{pmatrix}, \quad \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} \equiv \begin{pmatrix} 0.18 & 0.11 & 0.13 \\ 0.11 & 0.23 & 0.16 \\ 0.13 & 0.16 & 0.23 \end{pmatrix}, \tag{2.17}$$

and assume inequality and equality views as follows

$$f_{\boldsymbol{X}} \in \mathcal{C}_{\boldsymbol{X}} : \qquad \begin{cases} \mathbb{E}^{f_{\boldsymbol{X}}}\{X_1 - X_2\} \leq -0.01 \\ \mathbb{E}^{f_{\boldsymbol{X}}}\{X_2 - X_3\} \leq -0.01 \\ \mathbb{E}^{f_{\boldsymbol{X}}}\{X_3\} = 0.3. \end{cases} \tag{2.18}$$

Then the views are in the linear format as in (2.9), where

$$\boldsymbol{\zeta}_\mu \equiv \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix}, \quad \boldsymbol{\eta}^{view}_\mu \equiv \begin{pmatrix} -0.01 \\ -0.01 \\ 0.3 \end{pmatrix}. \tag{2.19}$$

Then, if we solve numerically the ensuing dual Lagrangian optimization (2.6)-(2.12), the updated expectation (2.10) reads

$$\bar{\boldsymbol{\mu}}_{\boldsymbol{X}} = \begin{pmatrix} 0.24 \\ 0.27 \\ 0.3 \end{pmatrix}, \tag{2.20}$$

which is consistent with the views since we have

$$\begin{cases} [\bar{\boldsymbol{\mu}}_{\boldsymbol{X}}]_1 - [\bar{\boldsymbol{\mu}}_{\boldsymbol{X}}]_2 = -0.02 \leq -0.01 \\ [\bar{\boldsymbol{\mu}}_{\boldsymbol{X}}]_2 - [\bar{\boldsymbol{\mu}}_{\boldsymbol{X}}]_3 = -0.03 \leq -0.01 \\ [\bar{\boldsymbol{\mu}}_{\boldsymbol{X}}]_3 = 0.3. \end{cases} \tag{2.21}$$

2.4 (In)equality views on first and second moments

In the more general case of (in)equality views on the first two moments as in (2.2), the dual Lagrangian problem (2.6) is an instance of a linearly constrained *convex* programming problem, since the dual Lagrangian $\mathcal{L}(\boldsymbol{t};\boldsymbol{\eta}^{view})$ (2.7) is a convex objective (30) in general.

In particular, it turns out that the $(\bar{k}_\mu + \bar{k}^2_\sigma) \times 1$ gradient vector explicitly reads [A.2.1]

$$\begin{aligned} \nabla_{\boldsymbol{t}}\mathcal{L}(\boldsymbol{t};\boldsymbol{\eta}^{view}) &\equiv \begin{pmatrix} \nabla_{\boldsymbol{t}_\mu}\mathcal{L}(\boldsymbol{t};\boldsymbol{\eta}^{view}) \\ \nabla_{\boldsymbol{t}_{\sigma,\sigma}}\mathcal{L}(\boldsymbol{t};\boldsymbol{\eta}^{view}) \end{pmatrix} \\ &= \begin{pmatrix} \boldsymbol{\zeta}_\mu & \boldsymbol{0}_{\bar{k}_\mu \times \bar{n}^2} \\ \boldsymbol{0}_{\bar{k}^2_\sigma \times \bar{n}} & \boldsymbol{\zeta}_{\sigma,\sigma} \end{pmatrix} \begin{pmatrix} \mu(\boldsymbol{t}) \\ vec(\sigma^2(\boldsymbol{t}) + \mu(\boldsymbol{t})\mu(\boldsymbol{t})') \end{pmatrix} - \begin{pmatrix} \boldsymbol{\eta}^{view}_\mu \\ vec(\boldsymbol{\eta}^{view}_{\sigma,\sigma}) \end{pmatrix}. \end{aligned} \tag{2.22}$$

where $(\mu(\boldsymbol{t}), \sigma^2(\boldsymbol{t}))$ are the multivariate functions as in (2.4)-(2.5); and $\boldsymbol{\zeta}_{\sigma,\sigma}$ is the $\bar{k}^2_\sigma \times \bar{n}^2$ matrix as in (1.8).

Similar to the above, the $(\bar{k}_\mu + \bar{k}^2_\sigma) \times (\bar{k}_\mu + \bar{k}^2_\sigma)$ Hessian matrix explicitly reads [A.2.1]

$$\begin{aligned} \nabla^2_{\boldsymbol{t},\boldsymbol{t}}\mathcal{L}(\boldsymbol{t};\boldsymbol{\eta}^{view}) &\equiv \begin{pmatrix} \nabla^2_{\boldsymbol{t}_\mu,\boldsymbol{t}_\mu}\mathcal{L}(\boldsymbol{t};\boldsymbol{\eta}^{view}) & \nabla^2_{\boldsymbol{t}_{\sigma},\boldsymbol{t}_\mu}\mathcal{L}(\boldsymbol{t};\boldsymbol{\eta}^{view})' \\ \nabla^2_{\boldsymbol{t}_{\sigma},\boldsymbol{t}_\mu}\mathcal{L}(\boldsymbol{t};\boldsymbol{\eta}^{view}) & \nabla^2_{\boldsymbol{t}_{\sigma,\sigma}}\mathcal{L}(\boldsymbol{t};\boldsymbol{\eta}^{view}) \end{pmatrix} \\ &= \begin{pmatrix} \boldsymbol{\zeta}_\mu & \boldsymbol{0}_{\bar{k}_\mu \times \bar{n}^2} \\ \boldsymbol{0}_{\bar{k}^2_\sigma \times \bar{n}} & \boldsymbol{\zeta}_{\sigma,\sigma} \end{pmatrix} \begin{pmatrix} \sigma^2(\boldsymbol{t}) & (\mu(\boldsymbol{t})' \otimes \sigma^2(\boldsymbol{t})) + (\sigma^2(\boldsymbol{t}) \otimes \mu(\boldsymbol{t})') \\ (\mu(\boldsymbol{t}) \otimes \sigma^2(\boldsymbol{t})) + (\sigma^2(\boldsymbol{t}) \otimes \mu(\boldsymbol{t})) & (\mathbb{I}_{\bar{n}^2} + \mathbb{K}_{\bar{n},\bar{n}})(\sigma^2(\boldsymbol{t}) \otimes \sigma^2(\boldsymbol{t})) + \epsilon(\boldsymbol{t}) \end{pmatrix} \begin{pmatrix} \boldsymbol{\zeta}_\mu & \boldsymbol{0}_{\bar{k}_\mu \times \bar{n}^2} \\ \boldsymbol{0}_{\bar{k}^2_\sigma \times \bar{n}} & \boldsymbol{\zeta}_{\sigma,\sigma} \end{pmatrix}', \end{aligned} \tag{2.23}$$

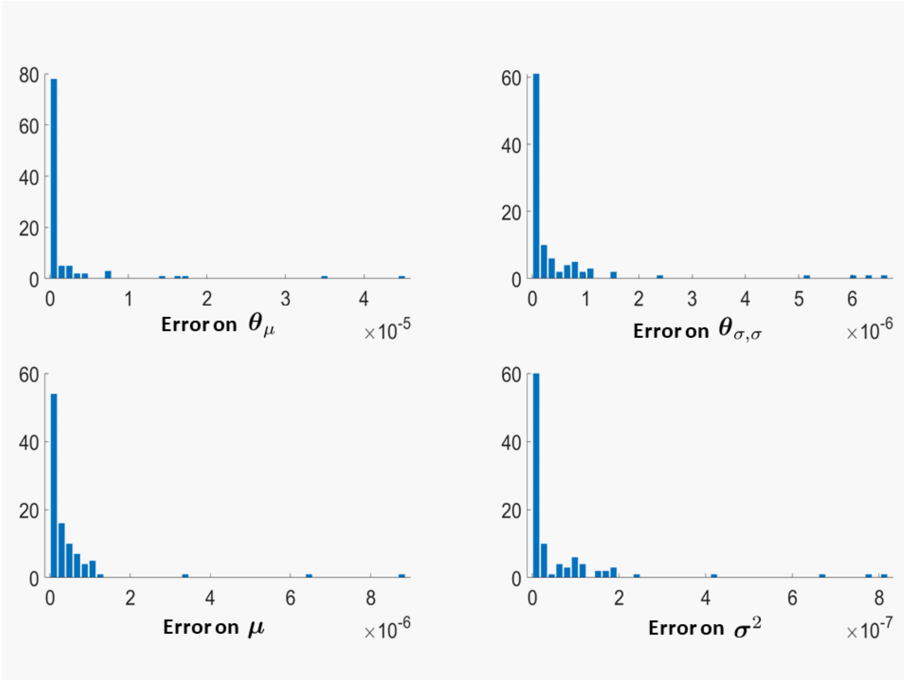

Figure 2.1: Approximation errors of the true MRE solution (Lagrange multipliers, expectation and covariance) via numerical implementation applied to 100 randomly chosen configurations of base parameters, equality views on first and second moments.

where $\mathbb{K}_{\bar{n},\bar{n}}$ denotes the $\bar{n}^2 \times \bar{n}^2$ commutation matrix [Magnus and Neudecker, 1979] and $\epsilon(\boldsymbol{t})$ is a multivariate function which we can write in terms of $(\mu(\boldsymbol{t}), \sigma^2(\boldsymbol{t}))$. Refer to [A.2.1] for more details.

According to the above, even in this case we can address numerically the dual Lagrangian problem via a built-in convex programming solver and we can further enhance its computation by feeding the *analytical* expressions of the first and second derivatives of the dual Lagrangian $\mathcal{L}(\boldsymbol{t}; \boldsymbol{\eta}^{view})$ (2.22)-(2.23) in the optimization algorithm. In Figure 2.1 we compare the numerical MRE solutions with the analytical counterparts in Chapter 1 for a random pool of normal base distributions (2.3).

Example 2.2. Consider $\bar{n} \equiv 6$ market variables $\boldsymbol{X} \equiv (X_1, \ldots, X_6)'$ with joint normal base distribution

$$\boldsymbol{X} \sim N(\underline{\boldsymbol{\mu}}_{\boldsymbol{X}}, \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}), \tag{2.24}$$

where

$$\underline{\boldsymbol{\mu}}_{\boldsymbol{X}} \equiv \begin{pmatrix} -1.14 \\ 0.10 \\ 0.72 \\ 2.58 \\ -0.66 \\ 0.18 \end{pmatrix}, \quad \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} \equiv \begin{pmatrix} 1 & 0.24 & 0.22 & -0.63 & 0.05 & -0.57 \\ \cdot & 1 & 0.33 & 0.42 & -0.63 & 0.28 \\ \cdot & \cdot & 1 & -0.29 & -0.79 & 0.03 \\ \cdot & \cdot & \cdot & 1 & -0.20 & 0.80 \\ \cdot & \cdot & \cdot & \cdot & 1 & -0.30 \\ \cdot & \cdot & \cdot & \cdot & \cdot & 1 \end{pmatrix}, \tag{2.25}$$

and assume inequality and equality views as follows

$$f_{\boldsymbol{X}} \in \mathcal{C}_{\boldsymbol{X}}: \quad \begin{cases} \mathbb{E}^{f_{\boldsymbol{X}}}\{\boldsymbol{\zeta}_{\mu}^{ineq}\boldsymbol{X}\} \leq \boldsymbol{\eta}_{\mu}^{view\_ineq} \\ \mathbb{E}^{f_{\boldsymbol{X}}}\{\boldsymbol{\zeta}_{\mu}^{eq}\boldsymbol{X}\} = \boldsymbol{\eta}_{\mu}^{view\_eq} \\ \mathbb{E}^{f_{\boldsymbol{X}}}\{\boldsymbol{\zeta}_{\sigma}\boldsymbol{X}\boldsymbol{X}'\boldsymbol{\zeta}_{\sigma}'\} \leq \boldsymbol{\eta}_{\sigma,\sigma}^{view}. \end{cases} \tag{2.26}$$

for specific combinations $\boldsymbol{\zeta}_{\mu}^{ineq}$, $\boldsymbol{\zeta}_{\mu}^{eq}$, $\boldsymbol{\zeta}_{\sigma}$; and features as follows

$$\boldsymbol{\eta}_{\mu}^{view\_ineq} \equiv \begin{pmatrix} 0.52 \\ -0.02 \\ -0.03 \end{pmatrix}, \quad \boldsymbol{\eta}_{\mu}^{view\_ineq} \equiv \begin{pmatrix} 0.30 \\ -1.25 \end{pmatrix} \tag{2.27}$$

$$\boldsymbol{\eta}_{\sigma,\sigma}^{view} \equiv \begin{pmatrix} 1.09 & -0.95 \\ -0.95 & 2.58 \end{pmatrix} \tag{2.28}$$

Then, if we solve numerically the ensuing dual Lagrangian optimization (2.6)-(2.12), the updated expectation (2.4) reads

$$\bar{\boldsymbol{\mu}}_{\boldsymbol{X}} = \begin{pmatrix} -0.87 \\ 0.23 \\ 0.20 \\ 2.79 \\ 0.00 \\ 0.39 \end{pmatrix}, \tag{2.29}$$

which is consistent with the views since we have

$$\begin{cases} \boldsymbol{\zeta}_{\mu}^{ineq}\bar{\boldsymbol{\mu}}_{\boldsymbol{X}} = \begin{pmatrix} 0.52 \\ -6.48 \\ -0.38 \end{pmatrix} \leq \begin{pmatrix} 0.52 \\ -0.02 \\ -0.03 \end{pmatrix} \\ \boldsymbol{\zeta}_{\mu}^{ineq}\bar{\boldsymbol{\mu}}_{\boldsymbol{X}} = \begin{pmatrix} 0.30 \\ -1.25 \end{pmatrix} \end{cases} \tag{2.30}$$

Similar to the above, the updated covariance (2.5) reads

$$\bar{\boldsymbol{\sigma}}_{\boldsymbol{X}}^2 = \begin{pmatrix} 1.66 & 0.52 & 0.52 & -0.78 & -0.15 & -0.33 \\ \cdot & 0.91 & 0.14 & 0.21 & -0.37 & 0.00 \\ \cdot & \cdot & 0.65 & -0.58 & -0.35 & -0.44 \\ \cdot & \cdot & \cdot & 0.92 & 0.09 & 0.48 \\ \cdot & \cdot & \cdot & \cdot & 0.48 & 0.25 \\ \cdot & \cdot & \cdot & \cdot & \cdot & 0.38 \end{pmatrix}, \tag{2.31}$$

which is consistent with the views since we have

$$\boldsymbol{\zeta}_{\sigma}(\bar{\boldsymbol{\sigma}}_{\boldsymbol{X}}^2 + \bar{\boldsymbol{\mu}}_{\boldsymbol{X}}\bar{\boldsymbol{\mu}}_{\boldsymbol{X}}')\boldsymbol{\zeta}_{\sigma}' = \begin{pmatrix} 1.09 & -0.95 \\ -0.95 & 2.48 \end{pmatrix} \leq \begin{pmatrix} 1.09 & -0.95 \\ -0.95 & 2.58 \end{pmatrix}. \tag{2.32}$$

2.5 Case study: ranking views

In this section we use MRE to build enhanced systematic strategies, optimally processing ranking (inequality) trading signals in the equity market. More precisely, following the most standard approach to this problem popularized by [Meucci et al., 2011], here we proceed at each generic time t by backtesting signal, as follows.

Step 1.We denote by $X_n \equiv \Pi_{n,t\rightarrow t+1}$ the next-step P&L of the n-th instrument, and we assume the joint (base) distribution to be normal as in (2.1), where the base expectation $\underline{\boldsymbol{\mu}}_{\boldsymbol{X}}$

and covariance $\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}$ are estimated from the past observations of the linear returns, say using exponentially weighted moving average.

Example 2.3. Let us consider a market based on $\bar{n} = 30$ instruments, say stocks or total return indexes, where the true dynamic for the dividend-adjusted values follows a simple geometric Brownian motion according to a standard Black-Merton-Scholes model [Black and Scholes, 1973].
This means that the log-values follow a random walk at discrete times

$$\ln V_{n,t+1} = \ln V_{n,t} + \varepsilon_{n,t\to t+1}, \tag{2.33}$$

with normal shocks, or compounded returns

$$\boldsymbol{\varepsilon}_{t\to t+1} \sim \text{i.i.d. } N(\boldsymbol{\mu}^*_{\boldsymbol{\varepsilon}}, \boldsymbol{\sigma}^{2*}_{\boldsymbol{\varepsilon}}). \tag{2.34}$$

According to this framework, we generate fake data from January 2002 to January 2017 with a weekly time step $t \to t+1$, so that linear returns behave similarly to compounded returns, and hence can be assumed normal in first approximation

$$R_{n,t\to t+1} \equiv \frac{V_{n,t+1} - V_{n,t}}{V_{n,t}} \approx \varepsilon_{n,t\to t+1} \equiv \ln(V_{n,t+1}/V_{n,t}). \tag{2.35}$$

We start collecting data in January 2012 estimate every week the expectation $\underline{\boldsymbol{\mu}}_{\boldsymbol{\varepsilon}}$ and covariance $\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{\varepsilon}}$ via historical approach, rolling on one year of data from the past series of linear returns, and deduce the ensuing estimate for the next-step P&L's

$$\boldsymbol{\Pi}_{t\to t+1} \sim N(\underline{\boldsymbol{\mu}}_{\boldsymbol{\Pi}}, \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{\Pi}}), \tag{2.36}$$

where

$$\underline{\boldsymbol{\mu}}_{\boldsymbol{\Pi}} = Diag(\boldsymbol{v}_t)\underline{\boldsymbol{\mu}}_{\boldsymbol{\varepsilon}}, \quad \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{\Pi}} = Diag(\boldsymbol{v}_t)\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{\varepsilon}}Diag(\boldsymbol{v}_t). \tag{2.37}$$

We repeat the procedure up to January 2017.

Step 2. We focus on an observable characteristic of a set of $\bar{n}$ instruments, which is deemed to have predictive power, say for instance, for stocks, a momentum/reversal indicator, or a value indicator such as the price/earnings ratio. Then we sort the $\bar{n}$ assets according to the value of the given characteristic. In our example, the stock $n = 1$ has the lowest momentum, the stock $n = 2$ has the second-lowest momentum, and so on, until the stock $n = \bar{n}$ has the highest momentum. The rationale of this step is that, if the signal is truly predictive, a lower ranking should give rise to a lower information ratios.

This is clearly a view in the following format

$$f_{\boldsymbol{X}} \in \mathcal{C}_{\boldsymbol{X}} : \qquad \frac{\mathbb{E}^{f_{\boldsymbol{X}}}\{X_1\}}{\mathbb{S}d^{f_{\boldsymbol{X}}}\{X_1\}} \leq \frac{\mathbb{E}^{f_{\boldsymbol{X}}}\{X_2\}}{\mathbb{S}d^{f_{\boldsymbol{X}}}\{X_2\}} \leq \cdots \leq \frac{\mathbb{E}^{f_{\boldsymbol{X}}}\{X_{\bar{n}}\}}{\mathbb{S}d^{f_{\boldsymbol{X}}}\{X_{\bar{n}}\}}, \tag{2.38}$$

where $X_n \equiv \Pi_{n,t\to t+1}$ denotes the next-step P&L of the n-th instrument.

Example 2.4. We continue from Example 2.3. In each week, we suppose to know the exact ranking of the Sharpe ratios behind the true P&L's distribution

$$n \to rank(n): \quad \frac{\mu^*_{\Pi;rank(n)}}{\sigma^*_{\Pi;rank(n)}} \leq \frac{\mu^*_{\Pi;rank(n+1)}}{\sigma^*_{\Pi;rank(n+1)}}, \tag{2.39}$$

which hence implies the views (2.38), up to re-ordering the stocks $n = 1, \ldots, \bar{n}$.

Step 3. The most common approach to address the views as in (2.38) popularized by [Grinold and Kahn, 1999], and later by [Park, 2010], [Wang and Kochard, 2011], [Moskowitz et al., 2012], [Asness et al., 2013a], [Asness et al., 2013b], [Menchero et al., 2013], updates the expectations, setting them to be proportional to their relative ranking and volatility, as follows

$$[\bar{\boldsymbol{\mu}}_{\boldsymbol{X}}]_n \equiv \eta\sqrt{[\bar{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}]_{n,n,}}(n - \frac{\bar{n}+1}{2}), \quad n = 1, \ldots, \bar{n}, \tag{2.40}$$

leaving the covariances the same as the base counterparts (2.11), or

$$\bar{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} \equiv \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}; \tag{2.41}$$

where the constant η is set as $2/(\bar{n}-1)$ so that the ex-ante Sharpe ratios are bounded as follows

$$-1 = \frac{[\bar{\boldsymbol{\mu}}_{\boldsymbol{X}}]_1}{\sqrt{[\bar{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}]_{1,1}}} \leq \cdots \leq \frac{[\bar{\boldsymbol{\mu}}_{\boldsymbol{X}}]_{\bar{n}}}{\sqrt{[\bar{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}]_{\bar{n},\bar{n}}}} = 1. \tag{2.42}$$

The above procedure presents several problems. First, the approach gives rise to ex-ante Sharpe ratios which are always bounded between -1 and 1 across time, which is a more strict restriction than what the views state (2.38). Second the expectation update (2.40) does not take into account the whole informations of the past data, as represented by the base expectations $\underline{\boldsymbol{\mu}}_{\boldsymbol{X}}$. Finally the approach does not change the volatilities (2.41), whereas the views (2.38) clearly also involves the volatilities. [Almgren and Chriss, 2006] provide an alternative approach to process inequality views. However, this implementation presents similar problems as in [Grinold and Kahn, 1999].

To address these issues, [Meucci et al., 2011] and [Meucci et al., 2014] propose a more enhanced solution via the so-called Factor Entropy Pooling (FEP) methodology. More precisely, the authors consider the relative entropy among normal distributions

$$\begin{aligned}\mathcal{E}(\boldsymbol{\mu}, \boldsymbol{\sigma}^2 \| \underline{\boldsymbol{\mu}}, \underline{\boldsymbol{\sigma}}^2) = \tfrac{1}{2}(&\mathrm{tr}(\boldsymbol{\sigma}^2(\underline{\boldsymbol{\sigma}}^2)^{-1}) - \ln|\boldsymbol{\sigma}^2(\underline{\boldsymbol{\sigma}}^2)^{-1}| \\ &+ (\boldsymbol{\mu} - \underline{\boldsymbol{\mu}})'(\underline{\boldsymbol{\sigma}}^2)^{-1}(\boldsymbol{\mu} - \underline{\boldsymbol{\mu}}) - \bar{n}),\end{aligned} \tag{2.43}$$

as distance between expectations and covariances and solve numerically the following minimization

$$(\bar{\boldsymbol{\mu}}_{\boldsymbol{X}}, \bar{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}) \equiv \underset{(\boldsymbol{\mu},\boldsymbol{\sigma}^2)\in\mathcal{C}}{\operatorname{argmin}}\, \mathcal{E}(\boldsymbol{\mu}, \boldsymbol{\sigma}^2 \| \underline{\boldsymbol{\mu}}_{\boldsymbol{X}}, \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}), \tag{2.44}$$

under the constraints implied by the views on the ex-ante Sharpe ratios (2.38)

$$\mathcal{C}: \qquad \frac{\mu_n}{\sigma_n} \leq \frac{\mu_{n+1}}{\sigma_{n+1}} - \eta, \quad n = 1, \ldots, \bar{n}-1. \tag{2.45}$$

Finally, in order to simplify the problem (2.44) and substantially reduce the large number of parameters $(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$ to optimize, they reformulate the covariances according to a "low-rank-diagonal" (factor) structure

$$\boldsymbol{\sigma}^2 \equiv \boldsymbol{b}\boldsymbol{b}' + Diag(\boldsymbol{d} \circ \boldsymbol{d}), \tag{2.46}$$

where $\boldsymbol{b}$ is an $\bar{n} \times \bar{h}$ matrix ($\bar{h} \ll \bar{n}$), and $\boldsymbol{d}$ is an $\bar{n} \times 1$ vector. However, this implementation is not consistent with the actual MRE solution (16), which is based on the dual Lagrangian optimization (27), as explained in details in Chapter 1. Moreover, the low-rank-diagonal parametrization is affected by identification problems: the parameters $\boldsymbol{b}$ identifies $\boldsymbol{\sigma}^2$ up to a $\bar{h} \times \bar{h}$ rotation matrix. Finally, the FEP implementation acts also on correlations, while the views (2.38) refer only to expectations and volatilities. Instead, it is more plausible that the minimal MRE distortion from the base, according to the views (2.38), does not involve correlations.

For the above reasons, here we propose to address the problem via MRE (2.6).

More precisely, we reformulate the ranking views (2.38) as follows

$$f_{\boldsymbol{X}} \in \mathcal{C}_{\boldsymbol{X}}: \qquad \begin{cases} \frac{\mathbb{E}^{f_{\boldsymbol{X}}}\{X_1\}}{\mathbb{S}d^{f_{\boldsymbol{X}}}\{X_1\}} \leq \frac{\mathbb{E}^{f_{\boldsymbol{X}}}\{X_2\}}{\mathbb{S}d^{f_{\boldsymbol{X}}}\{X_2\}} \leq \cdots \leq \frac{\mathbb{E}^{f_{\boldsymbol{X}}}\{X_{\bar{n}}\}}{\mathbb{S}d^{f_{\boldsymbol{X}}}\{X_{\bar{n}}\}} \\ \mathbb{C}v^{f_{\boldsymbol{X}}}\{\boldsymbol{X}\} = \bar{\boldsymbol{\sigma}}^{2view}_{\boldsymbol{X}}, \end{cases} \tag{2.47}$$

where we require $\bar{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}$ the updated covariance to not alter only the original base correlations

$$\bar{\boldsymbol{\sigma}}^{2view}_{\boldsymbol{X}} \equiv Diag(\bar{\boldsymbol{\sigma}}^{view}_{\boldsymbol{X};vol}) \times \underline{\boldsymbol{c}}^2_{\boldsymbol{X}} \times Diag(\bar{\boldsymbol{\sigma}}^{view}_{\boldsymbol{X};vol}), \tag{2.48}$$

where $\bar{\boldsymbol{\sigma}}^{view}_{\boldsymbol{X};vol}$ is a suitable $\bar{n} \times 1$ vector of new target volatilities, which we can calibrate or set exogenously.

Then, the reformulated views (2.47) are equivalent to inequality views on only expectations [A.2.4] as in (2.9), where $\boldsymbol{\zeta}_\mu$ is an $(\bar{n}-1) \times \bar{n}$ matrix defined as follows

$$\boldsymbol{\zeta}_\mu \equiv \begin{pmatrix} \frac{1}{[\bar{\boldsymbol{\sigma}}^{view}_{\boldsymbol{X};vol}]_1} & -\frac{1}{[\bar{\boldsymbol{\sigma}}^{view}_{\boldsymbol{X};vol}]_2} & 0 & \cdots & 0 \\ 0 & \frac{1}{[\bar{\boldsymbol{\sigma}}^{view}_{\boldsymbol{X};vol}]_2} & -\frac{1}{[\bar{\boldsymbol{\sigma}}^{view}_{\boldsymbol{X};vol}]_3} & \cdots & 0 \\ \cdot & \cdot & \ddots & \ddots & \cdot \\ 0 & 0 & \cdots & \frac{1}{[\bar{\boldsymbol{\sigma}}^{view}_{\boldsymbol{X};vol}]_{\bar{n}-1}} & -\frac{1}{[\bar{\boldsymbol{\sigma}}^{view}_{\boldsymbol{X};vol}]_{\bar{n}}} \end{pmatrix}, \tag{2.49}$$

and $\boldsymbol{\eta}_{\mu}^{view}$ is an $(\bar{n}-1)\times 1$ vector defined as follows

$$\boldsymbol{\eta}_{\mu}^{view} \equiv \begin{pmatrix} -\eta_1 \\ -\eta_2 \\ \cdot \\ -\eta_{\bar{n}} \end{pmatrix}, \tag{2.50}$$

where η_n are a positive scalar which we set to induce stronger inequalities in (2.47).

In this way we can compute numerically the updated expectation $\bar{\boldsymbol{\mu}}_{\boldsymbol{X}}$ as in (2.10), updating the covariance as in (2.48).

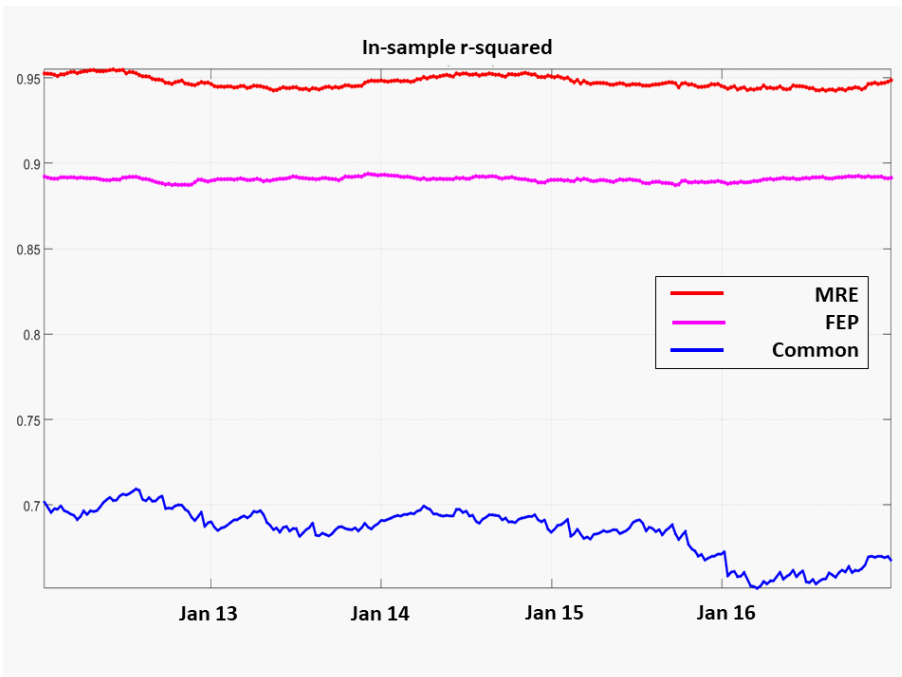

Figure 2.2: In-sample r-squared between the true P&L's distribution and updated distributions. Approaches: Minimum Relative Entropy (red); Factor Entropy Pooling (magenta); common approach (blue).

Example 2.5. We continue from Example 2.4 and we update parameters via Common approach (2.40); and FEP approach (2.44); and MRE via dual Lagrangian optimization (2.6)-(2.12), setting the inequality buffer in (2.50) as in the common approach (2.40), i.e. $\eta_n = 2/(\bar{n}-1)$ and the target volatilities $\bar{\boldsymbol{\sigma}}_{\boldsymbol{X};vol}^{view}$ as the base standard deviation.
We display in Figure 2.2 the in-sample r-squared from the true P&L's distribution, as measured by the relative entropy between the updated and true parameters $r^2 = 1 - \frac{1}{\bar{n}}\mathcal{E}(\bar{\boldsymbol{\mu}}_{\boldsymbol{X}}, \bar{\boldsymbol{\sigma}}_{\boldsymbol{X}}^2 \| \boldsymbol{\mu}_{\boldsymbol{X}}^*, \boldsymbol{\sigma}_{\boldsymbol{X}}^{2*})$.

Step 4. We construct an optimal portfolio, based on the updated expectation $\bar{\boldsymbol{\mu}}_{\boldsymbol{X}}$, such as maximum-expected long-short portfolio with constant target volatility, imposing constraints on the portfolio concentration similar to [Lobo et al., 2007]

$$\boldsymbol{h}^* \equiv \underset{\boldsymbol{h}\in\mathcal{C}}{\operatorname{argmax}}(\underbrace{\bar{\boldsymbol{\mu}}'_{\boldsymbol{X}}\boldsymbol{h}}_{\text{exp. P\&L}} - \underbrace{\boldsymbol{t}'\boldsymbol{h}}_{\text{trans. cost}}), \tag{2.51}$$

Refer [Meucci et al., 2011] for more details.

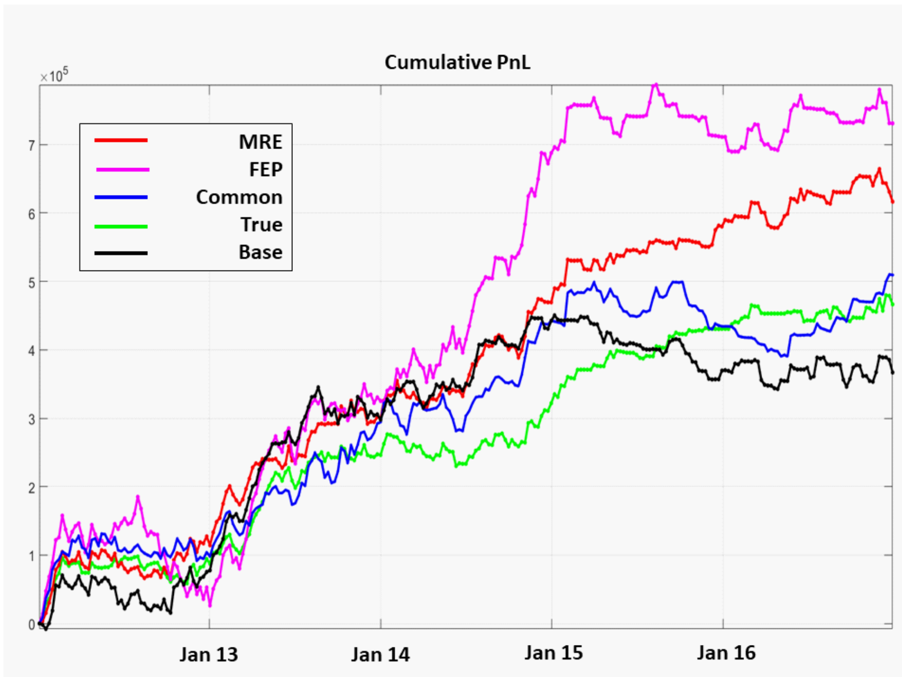

Figure 2.3: Cumulative P&L's of systematic strategies on Black-Scholes generated data. Approaches: Minimum Relative Entropy (red); Factor Entropy Pooling (magenta); common approach (blue); true (green); base (black).

Example 2.6. We continue from Example 2.5 and compute the optimal portfolio (2.51), where we set the transaction costs $\boldsymbol{t}$ as 5 basis points of the market value, and where we set the volatility target such that the dollar volatility is bounded at 10,000$.
In Figure 2.3 we display the cumulative P&L ensuing the optimal portfolios computed from the common approach (2.40)-(2.41), FEP approach (2.44)-(2.45) and the MRE approach for a total of 260 rebalancing dates. For comparison, we display the performance of the optimal portfolios stemming from the base P&L's distribution and true counterpart, which we consider as our benchmark.
In table below we report the weekly out-of-sample Sharpe ratio for the true optimal strategy,

and for the respective other strategies. In parenthesis we also report the p-value of the difference between the Sharpe ratio of each strategy from the benchmark, which is computed using the methodology by [Jobson and Korkie, 1981] and [DeMiguel et al., 2009].

Strategy ($\bar{n} = 30$)	Sharpe ratios	Mean	Std. Dev.
True	0.22	$10^3 \times 1.79$	$10^4 \times 0.79$
MRE	0.23 (0.80)	$10^3 \times 2.37$	$10^4 \times 0.99$
Common	0.18 (0.45)	$10^3 \times 1.95$	$10^4 \times 1.08$
FEP	0.19 (0.51)	$10^3 \times 2.81$	$10^4 \times 1.47$
Base (no views)	0.13 (0.17)	$10^3 \times 1.41$	$10^4 \times 1.02$

Table 2.1: Sharpe ratios and p-values of systematic strategies on Black-Scholes generated data

Example 2.7. To illustrate the backtesting strategy on real data, we consider a market of $\bar{n} = 30$ equities in the Dow Jones Index (constituents as of June 2012). For those equities, we consider weekly prices from January 2002 to January 2017. Within this framework we construct a predictive signal similarly to [Park, 2010]. More precisely, for each stock n, at the current time t, we define as "momentum" the quotient of a short term momentum and a long term standard deviation estimated by exponentially weighted moving average

$$mom_{n,t}^{\lambda,\gamma} \equiv \frac{\sum_{s\geq 0} e^{-\lambda s} r_{n,t-s}}{\sum_{s\geq 0} e^{-\lambda s}} \Big/ \sqrt{\frac{\sum_{s\geq 0} e^{-\gamma s} r_{n,t-s}^2}{\sum_{s\geq 0} e^{-\gamma s}}}. \tag{2.52}$$

In the above expression, values for the short-term decay coefficient λ correspond to a half-life of the order of a few days to a few weeks and typical values for the long-term decay coefficient γ correspond to a half-life of the order of a few weeks to a few months. Then, we reorder the stocks in such a way that $-mom_{1,t}^{\lambda,\gamma} \leq \cdots \leq -mom_{\bar{n},t}^{\lambda,\gamma}$, where the minus sign is set to implement a "reversal" strategy (plus sign for "momentum" strategy). The new ordering of stocks $n = 1, \ldots, \bar{n}$ implies the views (2.38). The backtest starts in January 2006 and portfolios are constructed every Wednesday for a total of 573 rebalancing dates similar to Example 2.6. In Figure 2.4 we display the cumulative P&L ensuing the optimal portfolios computed from the common approach (2.40)-(2.41), FEP approach (2.44)-(2.45) and the MRE approach. In order to visualize the predictive power of the signal (2.52) we also plot

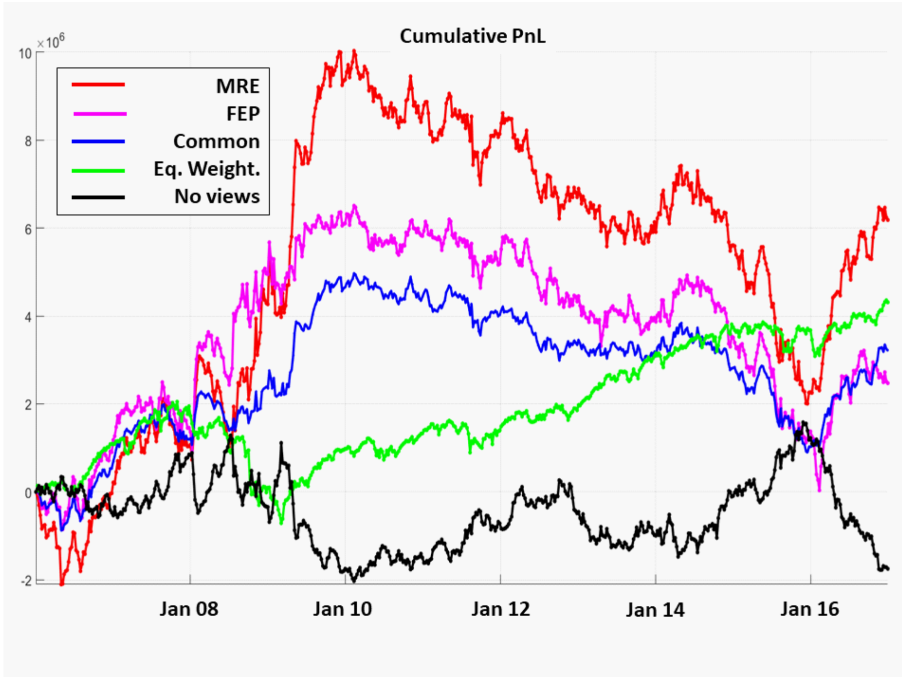

Figure 2.4: Cumulative P&L's of systematic strategies on Dow Jones index: Minimum Relative Entropy (red); Factor Entropy Pooling (magenta); common approach (blue); equally weighted (green); base (black).

the cumulative P&L of the optimal portfolios computed according to the base distribution (2.37), which hence do not take into account the views (2.38); and display the performance of an equally weighted portfolio which we consider as our benchmark. In table below we report the weekly out-of-sample Sharpe ratio for the true optimal strategy, and for the respective other strategies. In parenthesis we also report the p-value of the difference between the Sharpe ratios and p-values computed as in Table 2.1.

Strategy ($\bar{n} = 30$)	Sharpe ratios	Mean	Std. Dev.
Equally weighted	0.07	$10^3 \times 0.75$	$10^4 \times 1.02$
MRE	0.05 (0.63)	$10^3 \times 1.24$	$10^4 \times 2.64$
Common	0.04 (0.57)	$10^3 \times 0.55$	$10^4 \times 1.32$
FEP	0.01 (0.30)	$10^3 \times 0.40$	$10^4 \times 2.19$
Base (no views)	-0.01 (0.15)	$-10^3 \times 0.25$	$10^4 \times 1.48$

Table 2.2: Sharpe ratios and p-values of systematic strategies on Dow Jones data

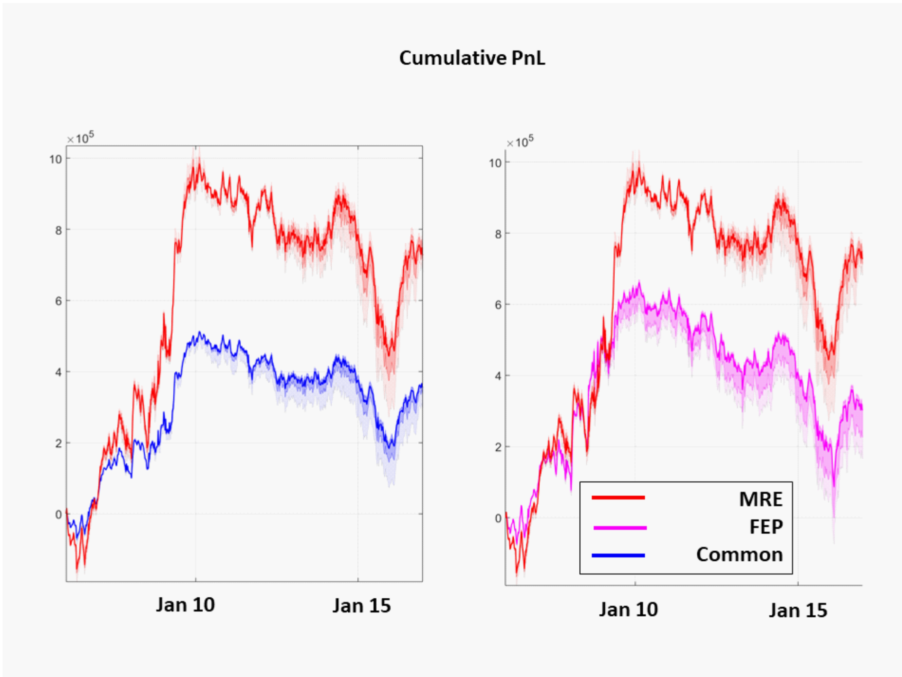

Figure 2.5: Fanplot of the backtested strategies on Dow Jones equity index using different decay parameters for the estimation of base expectations and base covariances, spanning a half-life from from 52 to 208 weeks. Approaches: Minimum Relative Entropy (red); Factor Entropy Pooling (magenta); common approach (blue).

Example 2.8. We continue from Example 2.7. For fairness of comparison among common, FEP and MRE approaches, we also performed the backtest with different values for the decay parameters in the base estimation, see Figure 2.5. The plot reports the median (solid line), the 50% percentile range (dim shading) and the 90% percentile range (dimmer shading) for each approach.
Finally, within this setup, we also plot the evolution of the respective Sharpe ratios and p-values computed as in Table 2.2, see Figure 2.6.

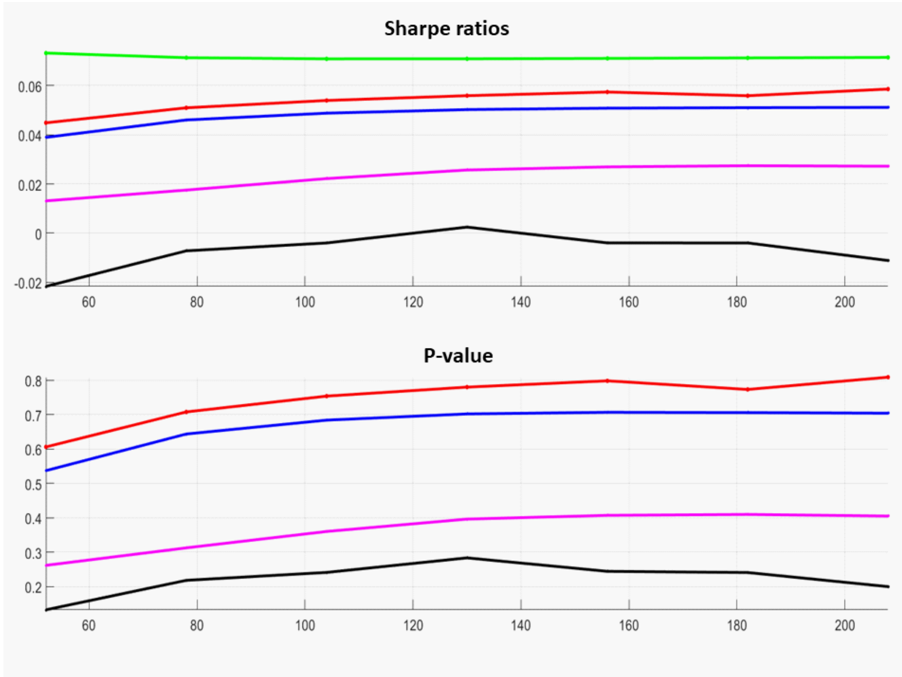

Figure 2.6: Sharpe ratios and p-values of a quantitative systematic strategy on Dow Jones equity index using different decay parameters for the estimation of base expectations and base covariances, spanning a half-life from from 52 to 208 weeks. Approaches: Minimum Relative Entropy (red); Factor Entropy Pooling (magenta); common approach (blue); equally weighted (green); base (black).

2.6 Conclusions

In this chapter we introduced a numerical approach for the parametric implementation of MRE under inequality views on the first two moments, generalizing results in Chapter 1. An interesting insight of portfolio theory is the formulation of dual Lagrangian problem in terms of mean-variance allocation. Other contributions are the analytical expressions of the gradient and the Hessian of the dual Lagrangian which allows to enhance the computational efficiency of the optimization algorithm.

Finally, we studied the effect of implementing the MRE into a systematic strategy based on ranking trade signals with respect to the former approaches introduced by [Grinold and Kahn, 1999] and [Meucci et al., 2011]. This comparison is undertaken using the Dow Jones empirical dataset from 2002 to 2017 as well as using simulated data. We found out that, as long as the signals are truly predictive, the in-sample errors in estimating means and covariances is significantly lower for the MRE implementation. Also we showed how the out-of-sample performance of MRE strategy, though lower than the equally weighted counterpart, tends to be definitely better than the classical implementation, which ignores the views; and higher with respect to the other approaches to views processing.

These results highlight the following facts. One, is the relevance of using not just empirical data but also other sources of informations, for instance cross-sectional characteristics, such as momentum/reversal indicators, or value indicators such as the price/earnings ratios. Two, when performing a particular signal-induced strategy for optimal allocation, the forecast of the right views plays a key role. Indeed, we stress that we are not proposing the MRE strategy as the best for all implementations, but only a consistent approach to suitably capture the signal informations, which can be used to test its predictive power in turn.

Chapter 3

A numerical implementation for fitting MRE models via stochastic approximation

3.1 Introduction

In problems for statistical modelling and inference, the principle of MRE have become popular through the years because of the relevant contributions by [Jaynes, 1957a], [Jaynes, 1957b]. The principle explains how to infer in our statistical models new informations that arise in the form of beliefs, or more precisely *views*. Such views are, in many cases, incomplete or *partial*, meaning that several distributional models for the variables of interest can take into account the views. In these circumstances, similar to Bayesian inference, the principle of MRE instructs us to choose the most plausible model displaying the minimal discrepancy, as measured by the relative entropy (17), from a prior knowledge, as represented by a reference (base) distribution (1), which is consistent with the views (16). See also [Cover and Thomas, 2006].

One cost imposed by this methodology is that the MRE solution (16) cannot be fitted analytically for general distributional models and arbitrary views, such as moment conditions on non-linear combinations of the variables. As a matter of fact, the parametric implementation of MRE, as introduced in Chapter 1 and 2, presents two barriers: the requirements of i) normality of the base distribution; and ii) views on the first two moments of linear combinations of the variables.

This implies the need to use numerical methods to fit the resulting MRE models in more general frameworks.

Most applications of MRE to date have involved numerical implementations via Monte Carlo sampling methods, such as stochastic approximation, or sample path optimization algorithms, see for instance [Schofield, 2007]. Among these techniques, the non-parametric implementation of MRE introduced in [Meucci, 2008] is very efficient, but typically inaccurate in particular when dealing with large dimensional markets.

Here we enhance and generalize [Meucci, 2008]. In particular, we show how through iterative Hamiltonian Monte Carlo iterative sampling [Chao et al., 2015], [Neal et al., 2011] we can fit the parameters of the MRE solution yielding a higher statistical significance than the non-parametric implementation. This can allow significant precision of the estimators especially when dealing with extreme scenarios.

The remainder of this chapter is organized as follows.

In Section 3.2 we introduce the MRE non-parametric framework for base distributions whose analytical expression is known modulo a normalizing constants and (in)equality views on arbitrary moments conditions, see Figure 1. In Section 3.3 we introduce the non-parametric implementation by [Meucci, 2008] and how to fit the MRE solution via Hamiltonian Monte Carlo simulations by [Chao et al., 2015], [Neal et al., 2011]. In Section 3.4 we show how to iterate the non-parametric procedure because of the invariance of the MRE solution under exponential tiltings and how to suitably set up stopping criteria. In Section 3.5 we compare the iterative approach with the non-parametric counterpart using the normal framework in Chapter 1 as benchmark. Finally, in Section 3.6 we draw some conclusions on the enhancements developed in this chapter.

Fully documented code is available on GitHub.

3.2 The model

Following the theoretical framework (4), here we address numerically the MRE problem (16) under no assumption on the base distribution and view function, see Figure 1.

More precisely, let us consider the following setup.

Suppose:
- a base distribution (1) whose analytical expression is perfectly known modulo a multiplicative constant term

$$\boldsymbol{X} \sim \underline{f}_{\boldsymbol{X}} : \qquad \underline{f}_{\boldsymbol{X}}(\boldsymbol{x}) \propto \underline{g}_{\boldsymbol{X}}(\boldsymbol{x}); \tag{3.1}$$

- (in)equality views on generalized expectation as in (22)

$$f_{\boldsymbol{X}} \in \mathcal{C}_{\boldsymbol{X}} : \qquad \mathbb{E}^{f_{\boldsymbol{X}}}\left\{\zeta^{view}(\boldsymbol{X})\right\} \leqq \boldsymbol{\eta}^{view}, \tag{3.2}$$

with an *arbitrary* view function ζ^{view}.

Then, the MRE updated distribution (23) must be an exponential twist of the base distribution (23)

$$\boldsymbol{X} \sim \bar{f}_{\boldsymbol{X}} : \qquad \bar{f}_{\boldsymbol{X}} \propto \bar{g}_{\boldsymbol{X}}, \tag{3.3}$$

where

$$\bar{g}_{\boldsymbol{X}}(\boldsymbol{x}) \equiv \underline{g}_{\boldsymbol{X}}(\boldsymbol{x}) e^{\boldsymbol{\theta}^{view\prime}\zeta^{view}(\boldsymbol{x})}, \tag{3.4}$$

and where the optimal Lagrange multipliers $\boldsymbol{\theta}^{view} \equiv (\theta_1^{view}, \ldots, \theta_{\bar{k}}^{view})'$ are the solutions of the following dual Lagrangian problem (27), which we report here

$$\boldsymbol{\theta}^{view} \equiv \underset{\boldsymbol{t} \leqq \boldsymbol{0}}{\operatorname{argmin}}\, \mathcal{L}(\boldsymbol{t}; \boldsymbol{\eta}^{view}). \tag{3.5}$$

Now our goal is to compute:
i) the optimal Lagrange multipliers $\boldsymbol{\theta}^{view}$ (3.5);
ii) the updated distribution $\bar{f}_{\boldsymbol{X}}$ (3.3).

Since in full generality we cannot proceed via analytical computations as in Chapter 1 and 2, here we focus on addressing the MRE problem via simulations. To this purpose we rely on approximations through scenario-probability distributions

$$f_{\boldsymbol{X}}(\boldsymbol{x}) \approx \textstyle\sum_{j=1}^{\bar{j}} p^{(j)} \delta^{(\boldsymbol{x}^{(j)})}(\boldsymbol{x}) \quad \Leftrightarrow \quad \{\mathcal{X}, \boldsymbol{p}\}, \tag{3.6}$$

which are fully identified by:
- an $\bar{n} \times \bar{j}$ panel-matrix of joint scenarios

$$\mathcal{X} \equiv \begin{pmatrix} x_1^{(1)} & \cdot & x_1^{(j)} & \cdot & x_1^{(\bar{j})} \\ \vdots & & \vdots & & \vdots \\ x_n^{(1)} & \cdot & x_n^{(j)} & & x_n^{(\bar{j})} \\ \vdots & & \vdots & & \vdots \\ x_{\bar{n}}^{(1)} & & x_{\bar{n}}^{(j)} & & x_{\bar{n}}^{(\bar{j})} \end{pmatrix}; \tag{3.7}$$

- a $\bar{j} \times 1$ vector of probabilities $\boldsymbol{p} \equiv (p^{(1)}, \ldots, p^{(\bar{j})})'$ which are positive and sum to one

$$\boldsymbol{p} \geq \boldsymbol{0}, \quad \textstyle\sum_{j=1}^{\bar{j}} p^{(j)} = 1. \tag{3.8}$$

See also [Meucci, 2019] for more details.

Then, we proceed as follows.

First, we generate via Hamiltonian Monte Carlo (HMC) simulations [Chao et al., 2015], [Neal et al., 2011] a scenario-probability distribution (3.6) that approximates the base distribution $\underline{f}_{\boldsymbol{X}}$ (3.1)

$$\{\underline{\mathcal{X}}, \underline{\boldsymbol{p}}\} \approx \underline{f}_{\boldsymbol{X}} \quad \stackrel{\text{HMC}}{\Leftarrow} \quad g_{\boldsymbol{X}}. \tag{3.9}$$

Next, according to the non-parametric MRE [Meucci, 2008], given $\{\underline{\mathcal{X}}, \underline{\boldsymbol{p}}\}$ we can approximate both i) the optimal Lagrange multipliers $\boldsymbol{\theta}^{view}$ (3.5); and ii) the updated distribution $\bar{f}_{\boldsymbol{X}}$ (3.3)

$$\left.\begin{array}{l} \hat{\boldsymbol{\theta}}^{view} \approx \boldsymbol{\theta}^{view} \\ \{\underline{\mathcal{X}}, \bar{\boldsymbol{p}}\} \approx \bar{f}_{\boldsymbol{X}} \end{array}\right\} \quad \Leftarrow \quad \left\{\begin{array}{l} \{\underline{\mathcal{X}}, \underline{\boldsymbol{p}}\} \approx \underline{f}_{\boldsymbol{X}} \\ \{\zeta^{view}, \boldsymbol{\eta}^{view}\} \end{array}\right. . \tag{3.10}$$

However, the statistical approximation of the scenario-probability distribution (3.6) represented by $\{\underline{\mathcal{X}}, \bar{\boldsymbol{p}}\}$ can be significantly poor due to the curse of dimensionality. Moreover, as we proceed to show, the approximation error for the Lagrange multipliers $\boldsymbol{\theta}^{view}$ can be significantly large, in particular when we deal with extreme views.

To solve these issues, here we propose to iterate the non-parametric MRE:
1) perform the steps (3.9) and (3.10);

2) compute the analytical approximation for the generator $\bar{g}_{\boldsymbol{X}}$ of the updated distribution (3.4)

$$\hat{g}_{\boldsymbol{X}}(\boldsymbol{x}) \leftarrow \underline{g}_{\boldsymbol{X}}(\boldsymbol{x}) e^{\hat{\boldsymbol{\theta}}^{view\prime} \zeta^{view}(\boldsymbol{x})}; \tag{3.11}$$

3) if convergence criteria are satisfied, output the fitted generator and the updated distribution

$$\hat{g}_{\boldsymbol{X}}, \{\underline{\mathcal{X}}, \bar{\boldsymbol{p}}\} \approx \bar{f}_{\boldsymbol{X}}, \tag{3.12}$$

otherwise update the base generator (3.1)

$$\underline{g}_{\boldsymbol{X}}(\boldsymbol{x}) \leftarrow \hat{g}_{\boldsymbol{X}}(\boldsymbol{x}), \tag{3.13}$$

and go to 1).

3.3 Non-parametric MRE

Here we see step by step how to build the outputs $\underline{\mathcal{X}}$, $\hat{\boldsymbol{\theta}}^{view}$ and $\bar{\boldsymbol{p}}$ for the non-parametric MRE (3.10).

3.3.1 HMC sampling

In order to address the dual Lagrangian optimization (3.5), we start from the computation of the dual Lagrangian $\mathcal{L}(\boldsymbol{t}; \boldsymbol{\eta}^{view})$ (27), or the log-partition function $\psi(\boldsymbol{t})$ (25), which is a functional of the base distribution [A.3.1]

$$\psi(\boldsymbol{t}) = \psi[\underline{f}_{\boldsymbol{X}}](\boldsymbol{t}) \equiv \ln \underline{\mathbb{E}}\{e^{\boldsymbol{t}' \zeta^{view}(\boldsymbol{X})}\}, \tag{3.14}$$

where $\underline{\mathbb{E}}\{\cdot\}$ denotes the expectation under the base distribution $\underline{f}_{\boldsymbol{X}}$ (1). Hence, in the first step (3.9)-(3.10) we look for solving the expectation in (3.14) via Monte Carlo integration.

More precisely, here we rely on the so called **Hamiltonian Monte Carlo** (HMC) sampling approach [Chao et al., 2015], [Neal et al., 2011], which is a Markov chain Monte Carlo (MCMC) method that is more efficient than the Metropolis-Hastings algorithm [Berg, 2004]. Refer also to [Chib and Greenberg, 1995] and [Geweke, 1999] for more details.

As a matter of fact, the core feature of MCMC implementations, including HMC, is that they are "unaffected by scaling", i.e. they allow to sample from an arbitrary distribution $f_{\boldsymbol{X}}$ of the form

$$f_{\boldsymbol{X}}(\boldsymbol{x}) \equiv \gamma \times g_{\boldsymbol{X}}(\boldsymbol{x}), \tag{3.15}$$

with the only knowledge of $g(\boldsymbol{x})$. This is particularly useful to sample from an exponential family distribution (24)

$$f_{\boldsymbol{X}} \propto \underline{g}_{\boldsymbol{X}}(\boldsymbol{x}) e^{\boldsymbol{t}' \zeta^{view}(\boldsymbol{x})}, \tag{3.16}$$

for a given $\boldsymbol{t}$, including hence both base distribution (3.1) (case $\boldsymbol{t} = \boldsymbol{0}$) and updated counterpart (3.3) (case $\boldsymbol{t} = \boldsymbol{\theta}^{view}$).

Then, the HMC algorithm needs two inputs:

1. the log-pdf modulo constant terms

$$u(\boldsymbol{x}) = \boldsymbol{t}'\zeta^{view}(\boldsymbol{x}) + \ln \underline{f}_{\boldsymbol{X}}(\boldsymbol{x}); \tag{3.17}$$

2. (optionally) the respective gradient, which reads [A.3.3]

$$\nabla_{\boldsymbol{x}} u(\boldsymbol{x}) = \frac{1}{\underline{f}_{\boldsymbol{X}}(\boldsymbol{x})} \nabla_{\boldsymbol{x}} \underline{f}_{\boldsymbol{X}}(\boldsymbol{x}) + J_{\zeta^{view}}(\boldsymbol{x})'\boldsymbol{t}, \tag{3.18}$$

   and where $J_{\zeta^{view}}(\boldsymbol{x})$ denotes the $\bar{k} \times \bar{n}$ Jacobian matrix of the view function $\zeta^{view}(\boldsymbol{x})$.

According to the above, we can safely sample a large number $\bar{j}$ of simulations from the base distribution $\underline{f}_{\boldsymbol{X}} \propto \underline{g}_{\boldsymbol{X}}$ (3.1)

$$\underline{\mathcal{X}} \equiv HMC.sampler(\underline{g}_{\boldsymbol{X}}, \bar{j}), \tag{3.19}$$

and set uniform probabilities $\underline{\boldsymbol{p}} \equiv (\underline{p}^{(1)}, \ldots, \underline{p}^{(\bar{j})})'$

$$\underline{p}^{(j)} \equiv \frac{1}{\bar{j}}, \quad j = 1, \ldots, \bar{j}, \tag{3.20}$$

so that we can approximate the original base $\underline{f}_{\boldsymbol{X}}$ with its sample counterpart, because of the law of large numbers (LLN)

$$\underline{f}_{\boldsymbol{X}} \approx \hat{f}_{\boldsymbol{X}} \equiv \textstyle\sum_{j=1}^{\bar{j}} \underline{p}^{(j)} \delta^{(\underline{\boldsymbol{x}}^{(j)})}(\boldsymbol{x}) \quad \Leftrightarrow \quad \{\underline{\mathcal{X}}, \underline{\boldsymbol{p}}\}. \tag{3.21}$$

This allows also to approximate the log-partition function $\psi(\boldsymbol{t})$ with its sample counterpart

$$\psi[\underline{f}_{\boldsymbol{X}}](\boldsymbol{t}) \approx \psi[\hat{f}_{\boldsymbol{X}}](\boldsymbol{t}) = \hat{\psi}(\boldsymbol{t}; \{\{\underline{\mathcal{X}}, \underline{\boldsymbol{p}}\}) \equiv \ln(\textstyle\sum_{j=1}^{\bar{j}} \underline{p}^{(j)} e^{\boldsymbol{t}'\zeta^{view}(\underline{\boldsymbol{x}}^{(j)})}). \tag{3.22}$$

3.3.2 Lagrange multipliers fit

Once we approximated the log-partition function $\psi(\boldsymbol{t})$ (3.22), we proceed with the approximation of the optimal Lagrange multipliers $\boldsymbol{\theta}^{view}$ (3.5).

More precisely, we estimate $\boldsymbol{\theta}^{view}$ through the solution of the sample counterpart of the dual Lagrangian problem (3.5)

$$\hat{\boldsymbol{\theta}}^{view} \equiv \underset{\boldsymbol{t} \leqq \boldsymbol{0}}{\operatorname{argmin}}\, \hat{\psi}(\boldsymbol{t}; \{\underline{\mathcal{X}}, \underline{\boldsymbol{p}}\}) - \boldsymbol{t}'\boldsymbol{\eta}^{view}. \tag{3.23}$$

and fit in turn the updated distribution (3.3) through the ensuing exponential family distribution, or generator

$$\hat{g}_{\boldsymbol{X}}(\boldsymbol{x}) \equiv \underline{g}_{\boldsymbol{X}}(\boldsymbol{x}) e^{\hat{\boldsymbol{\theta}}^{view\prime}\zeta^{view}(\boldsymbol{x})} \approx \bar{g}_{\boldsymbol{X}}(\boldsymbol{x}). \tag{3.24}$$

Note that, similar to its theoretical counterpart (27), the optimization (3.23) is a low-dimensional convex programming problem and as such can be performed numerically via standard built-in solvers for minimization problems.

Moreover, the numerical computation can be further enhanced using the analytical expression of the gradient and Hessian of the sample dual Lagrangian $\hat{\mathcal{L}}(\boldsymbol{t};\boldsymbol{\eta}^{view})$ [A.3.2]. Refer also to [Meucci, 2008], [Kleywegt and Shapiro, 2001] and [Schofield, 2007] for more details.

Example 3.1. Consider $\bar{n} \equiv 2$ market variables $\boldsymbol{X} \equiv (X_1, X_2)'$ with joint normal base distribution

$$\boldsymbol{X} \sim N(\underline{\boldsymbol{\mu}}_{\boldsymbol{X}}, \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}), \tag{3.25}$$

where

$$\underline{\boldsymbol{\mu}}_{\boldsymbol{X}} \equiv \begin{pmatrix} 0.26 \\ 0.29 \end{pmatrix}, \quad \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} \equiv \begin{pmatrix} 0.18 & 0.11 \\ 0.11 & 0.23 \end{pmatrix}, \tag{3.26}$$

and suppose $\bar{k} = 1$ views on linear combinations of expectations as in (1.19), where

$$\boldsymbol{\zeta} \equiv \begin{pmatrix} 1 & -1 \end{pmatrix}, \quad \eta^{view} \equiv 1.02. \tag{3.27}$$

In this special case the true optimal Lagrange multipliers (3.3) can be computed analytically and reads (1.20)

$$\theta^{view*} = 5.53. \tag{3.28}$$

From the other hand, if we approximate the optimal Lagrange multipliers as in (3.23) we obtain

$$\hat{\theta}^{view} = 5.59. \tag{3.29}$$

3.3.3 Probabilities update

Once fitted the optimal Lagrange multipliers as in (3.23), the final step is to approximate the updated distribution $\bar{f}_{\boldsymbol{X}}$ (3.3), since the fitted distribution as in (3.24) is not analytically tractable in practice for computing statistical features, such as expectations, volatilities, quantiles etc.

Then we proceed as follows.

First, given the i.i.d. simulations $\underline{\mathcal{X}}$ stemming from the base distribution in (3.19), we arrange the base scenarios for the view variables $\boldsymbol{Z} \equiv \zeta^{view}(\boldsymbol{X})$ (2) into a $\bar{k} \times \bar{j}$ panel matrix

$$\underline{\mathcal{Z}} \equiv \begin{pmatrix} \zeta_1^{view}(\underline{\boldsymbol{x}}^{(1)}) & \cdot & \zeta_1^{view}(\underline{\boldsymbol{x}}^{(j)}) & \cdot & \zeta_1^{view}(\underline{\boldsymbol{x}}^{(\bar{j})}) \\ \vdots & & \vdots & & \vdots \\ \zeta_k^{view}(\underline{\boldsymbol{x}}^{(1)}) & \cdot & \zeta_k^{view}(\underline{\boldsymbol{x}}^{(j)}) & & \zeta_k^{view}(\underline{\boldsymbol{x}}^{(\bar{j})}) \\ \vdots & & \vdots & & \vdots \\ \zeta_{\bar{k}}^{view}(\underline{\boldsymbol{x}}^{(1)}) & & \zeta_{\bar{k}}^{view}(\underline{\boldsymbol{x}}^{(j)}) & & \zeta_{\bar{k}}^{view}(\underline{\boldsymbol{x}}^{(\bar{j})}) \end{pmatrix}. \tag{3.30}$$

Next, we estimate $\bar{f}_{\boldsymbol{X}}$ through a suitable scenario-probability distribution (3.6).

$$\hat{f}_{\boldsymbol{X}}(\boldsymbol{x}) \equiv \textstyle\sum_{j=1}^{\bar{j}} \bar{p}^{(j)} \delta^{(\underline{\boldsymbol{x}}^{(j)})}(\boldsymbol{x}) \quad \Leftrightarrow \quad \{\underline{\mathcal{X}}, \bar{\boldsymbol{p}}\}, \tag{3.31}$$

where the probabilities $\bar{\boldsymbol{p}}$ are positive weights which sum to one, defined via softmax function

$$\begin{aligned}\bar{\boldsymbol{p}} &= softmax(\ln(\underline{\boldsymbol{p}}) + \hat{\boldsymbol{\theta}}^{view\prime}\underline{\mathcal{Z}}) \\ &\equiv \frac{e^{\ln(\underline{\boldsymbol{p}})+\hat{\boldsymbol{\theta}}^{view\prime}\underline{\mathcal{Z}}}}{\sum_{j=1}^{\bar{j}} \underline{p}^{(j)} e^{\hat{\boldsymbol{\theta}}^{view\prime}[\underline{\mathcal{Z}}]_{\cdot,j}}},\end{aligned} \tag{3.32}$$

similar to the theoretical counterpart (24). Indeed note that the sample log-partition function (3.22) reads

$$\hat{\psi}(\hat{\boldsymbol{\theta}}^{view}; \{\underline{\mathcal{X}}, \underline{\boldsymbol{p}}\}) = \ln(\textstyle\sum_{j=1}^{\bar{j}} \underline{p}^{(j)} e^{\hat{\boldsymbol{\theta}}^{view\prime}[\underline{\mathcal{Z}}]_{\cdot,j}}). \tag{3.33}$$

To summarize, the output of the non-parametric approach is given by a scenario-probability set $\{\underline{\mathcal{X}}, \bar{\boldsymbol{p}}\}$ with *same old (base) scenarios* and *new (updated) probabilities* (3.32).

In particular, the non-parametric updated distribution (3.31) allows to approximate any statistical feature of the updated distribution, such as its expectation and covariance, simply through the sample counterparts stemming from $\{\underline{\mathcal{X}}, \bar{\boldsymbol{p}}\}$. In particular, it is easy to verify that under the non-parametric updated distribution (3.31) the view variables $\boldsymbol{Z} \equiv \zeta^{view}(\boldsymbol{X})$ satisfy the views (3.2), which means that the non-parametric mean of the view variables is constant with respect to the base scenarios $\underline{\mathcal{X}}$

$$\mathbb{E}^{\hat{f}_{\boldsymbol{X}}}\left\{\zeta^{view}(\boldsymbol{X})\right\} = \bar{\boldsymbol{p}}'\underline{\mathcal{Z}} = \boldsymbol{\eta}^{view}. \tag{3.34}$$

Example 3.2. We continue from Example 3.1. Since we consider views on expectations as in (1.19), we already know that the true updated distribution is normally distributed (1.15)

$$\boldsymbol{X} \sim N(\bar{\boldsymbol{\mu}}_{\boldsymbol{X}}, \bar{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}), \tag{3.35}$$

and in this case the updated expectation (1.21) and covariance (1.23) read

$$\bar{\boldsymbol{\mu}}_{\boldsymbol{X}} \equiv \left(\begin{smallmatrix} 0.65 \\ -0.37 \end{smallmatrix}\right), \quad \bar{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} \equiv \left(\begin{smallmatrix} 0.18 & 0.11 \\ 0.11 & 0.23 \end{smallmatrix}\right) \tag{3.36}$$

Instead, if we compute the updated probabilities (3.32), then the scenario-probability expectation and covariance are different

$$\hat{\boldsymbol{\mu}}_{\boldsymbol{X}} = \left(\begin{smallmatrix} 0.66 \\ -0.36 \end{smallmatrix}\right), \quad \hat{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} = \left(\begin{smallmatrix} 0.18 & 0.10 \\ 0.10 & 0.21 \end{smallmatrix}\right), \tag{3.37}$$

though the views are still satisfied

$$\boldsymbol{\zeta}\hat{\boldsymbol{\mu}}_{\boldsymbol{X}} = 1.02. \tag{3.38}$$

Moreover, also the whole non-parametric distribution significantly differs from the true updated distribution, see Figure 3.1.

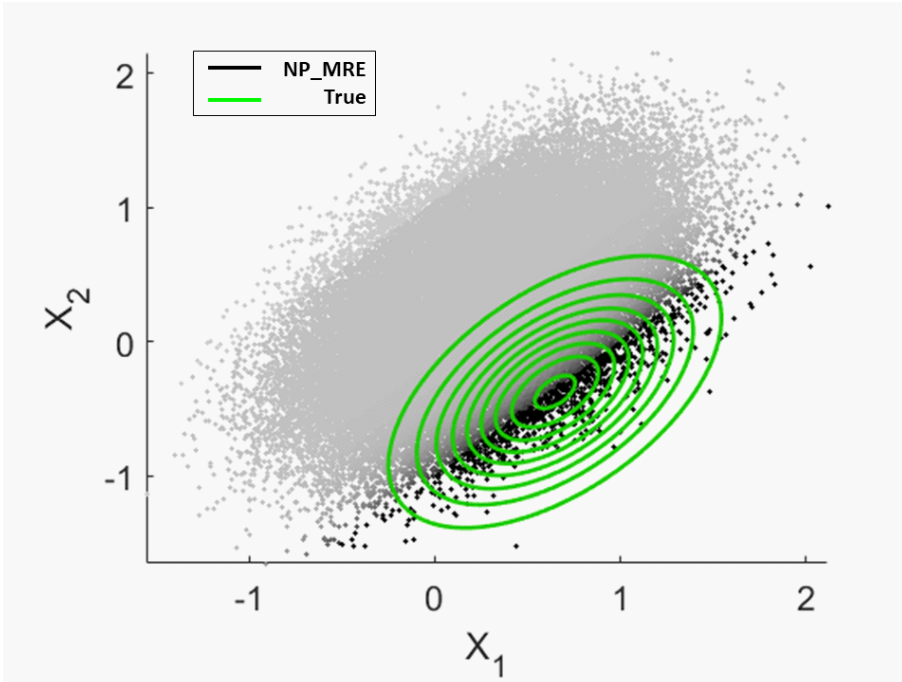

Figure 3.1: Comparison between true and non-parametric updated distributions.

3.4 Iterative MRE

In Section 3.3 we illustrated how to solve the MRE problem (16) for: i) a base distribution as in (1); and ii) (in)equality views on expectation (3.2) via the non-parametric MRE (3.10), as summarized in the following table.

$(\{\underline{\mathcal{X}}, \bar{\boldsymbol{p}}\}, \hat{\boldsymbol{\theta}}^{view}) = Scenarios.MRE(\zeta^{view}, \boldsymbol{\eta}^{view}, \underline{g}_{\boldsymbol{X}}, \bar{j})$	
1. Sample base scenarios	$\{\underline{\mathcal{X}}, \underline{\boldsymbol{p}}\} \overset{\text{HMC}}{\Leftarrow} (\underline{g}_{\boldsymbol{X}}, \bar{j})$ (3.19)
2. Compute Lagr. mult.	$\hat{\boldsymbol{\theta}}^{view} \Leftarrow (\zeta^{view}, \boldsymbol{\eta}^{view}, \{\underline{\mathcal{X}}, \underline{\boldsymbol{p}}\})$ (3.23)
3. Compute prob.	$\bar{\boldsymbol{p}} \Leftarrow (\hat{\boldsymbol{\theta}}^{view}, \zeta^{view}, \{\underline{\mathcal{X}}, \underline{\boldsymbol{p}}\})$ (3.32)

Table 3.1: Non-parametric MRE algorithm

Now we see the details behind the iterative approach described in steps from (3.11) to (3.13).

3.4.1 Invariance of the updated distribution under iteration

Let us denote by $\boldsymbol{\theta}^{view(1)} \equiv \hat{\boldsymbol{\theta}}^{view}$ the first-step Lagrange multipliers we approximated via non-parametric MRE (3.1).

It turns out that, as long as the views (3.2), which are identified by the function ζ^{view} and features $\boldsymbol{\eta}^{view}$, are the *same*, we can always consider the ensuing fitted exponential family distribution (3.11) as a new base distribution (3.1)

$$\underline{f}_{\boldsymbol{X}}(\boldsymbol{x}) \leftarrow f^{(1)}_{\boldsymbol{X}}(\boldsymbol{x}) \propto g^{(1)}_{\boldsymbol{X}}(\boldsymbol{x}) \equiv \underline{g}_{\boldsymbol{X}}(\boldsymbol{x}) e^{\boldsymbol{\theta}^{view(1)\prime}\zeta^{view}(\boldsymbol{x})}, \tag{3.39}$$

to input again in the non-parametric MRE routine (3.1). Then, the output of this first iteration is a new exponential family distribution (3.12) steered by a *new* vector of Lagrange multipliers $\boldsymbol{\epsilon}^{view(2)}$

$$f^{(2)}_{\boldsymbol{X}}(\boldsymbol{x}) \propto g^{(2)}_{\boldsymbol{X}}(\boldsymbol{x}) \equiv g^{(1)}_{\boldsymbol{X}}(\boldsymbol{x}) e^{\boldsymbol{\epsilon}^{view(2)\prime}\zeta^{view}(\boldsymbol{x})}. \tag{3.40}$$

The above procedure is perfectly consistent, in that it does not alter the distributional form of true updated distribution (3.3) we are looking for, because of the properties of the exponential family distributions.

As a matter of fact, $f^{(2)}_{\boldsymbol{X}}$ (3.40) is still an exponential family distribution under the original base $\underline{f}_{\boldsymbol{X}}$ (23), or [A.3.7]

$$g^{(2)}_{\boldsymbol{X}}(\boldsymbol{x}) = \underline{g}_{\boldsymbol{X}}(\boldsymbol{x}) e^{\boldsymbol{\theta}^{view(2)\prime}\zeta^{view}(\boldsymbol{x})}, \tag{3.41}$$

where the new Lagrange multipliers $\boldsymbol{\theta}^{view(2)}$ split into the sum of the new one $\boldsymbol{\epsilon}^{view(2)}$ and older one $\boldsymbol{\theta}^{view(1)}$ (3.23)

$$\boldsymbol{\theta}^{view(2)} \equiv \boldsymbol{\theta}^{view(1)} + \boldsymbol{\epsilon}^{view(2)}. \tag{3.42}$$

This automatically implies that, if we consider $f^{(1)}_{\boldsymbol{X}}$ as our new base distribution (3.39), then the true MRE updated distribution $\bar{f}_{\boldsymbol{X}}$ (3.3) must be the same [A.3.8]

$$\bar{f}_{\boldsymbol{X}} \equiv \underset{f_{\boldsymbol{X}} \in \mathcal{C}_{\boldsymbol{X}}}{\operatorname{argmin}} \mathcal{E}(f_{\boldsymbol{X}} \| \underline{f}_{\boldsymbol{X}}) = \underset{f_{\boldsymbol{X}} \in \mathcal{C}_{\boldsymbol{X}}}{\operatorname{argmin}} \mathcal{E}(f_{\boldsymbol{X}} \| f^{(1)}_{\boldsymbol{X}}). \tag{3.43}$$

Hence the MRE updated distribution $\bar{f}_{\boldsymbol{X}}$ is invariant under exponential tiltings of the base distribution. This means that we can iteratively repeat the non-parametric MRE routine (3.1) yielding each time a new exponential family distribution $f^{(i)}_{\boldsymbol{X}}$ as in (3.40)

$$f^{(i)}_{\boldsymbol{X}}(\boldsymbol{x}) \propto g^{(i)}_{\boldsymbol{X}}(\boldsymbol{x}) \equiv g^{(i-1)}_{\boldsymbol{X}}(\boldsymbol{x}) e^{\boldsymbol{\epsilon}^{view(i)\prime}\zeta^{view}(\boldsymbol{x})} = \underline{g}_{\boldsymbol{X}}(\boldsymbol{x}) e^{\boldsymbol{\theta}^{view(i)\prime}\zeta^{view}(\boldsymbol{x})}, \tag{3.44}$$

where

$$\boldsymbol{\theta}^{view(i)} \equiv \boldsymbol{\theta}^{view(i-1)} + \boldsymbol{\epsilon}^{view(i)}, \tag{3.45}$$

that is always a new potential candidate approximating the true unknown MRE updated distribution $\bar{f}_{\boldsymbol{X}}$.

In particular, it turns out that the sequence of distributions $\bar{f}^{(i)}_{\boldsymbol{X}}$ converges to the true updated distribution [A.3.9]

$$f^{(i)}_{\boldsymbol{X}} \underset{i\rightarrow\infty}{\rightarrow} \bar{f}_{\boldsymbol{X}} \quad \Leftrightarrow \quad \mathcal{E}(\bar{f}_{\boldsymbol{X}} || f^{(i)}_{\boldsymbol{X}}) \underset{i\rightarrow\infty}{\rightarrow} 0, \tag{3.46}$$

if the next-step Lagrange multipliers $\boldsymbol{\epsilon}^{view(i)}$ enough quickly converge to zero

$$\boldsymbol{\epsilon}^{view(i)} \underset{i\rightarrow\infty}{\rightarrow} \mathbf{0}. \tag{3.47}$$

Then, if the above holds the Lagrange multipliers $\boldsymbol{\theta}^{view(i)} = \sum_{l=1}^{i} \boldsymbol{\epsilon}^{view(l)}$ converge in turn to the true counterpart

$$\boldsymbol{\theta}^{view(i)} \underset{i\to\infty}{\longrightarrow} \boldsymbol{\theta}^{view}. \tag{3.48}$$

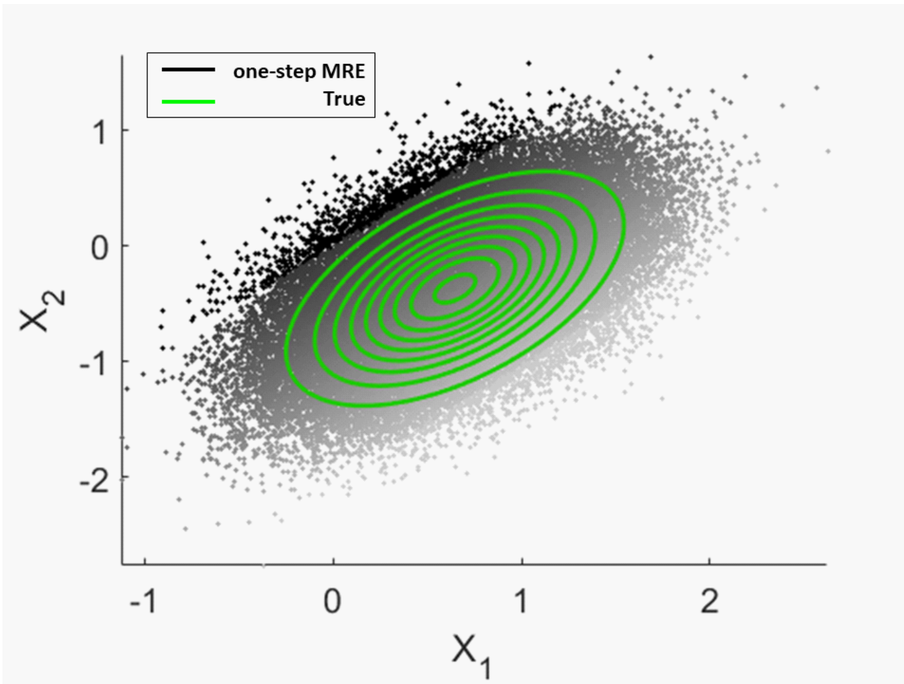

Figure 3.2: Comparison between true and one-step scenario-probability distributions.

Example 3.3. We continue from Example 3.2. If we perform one iteration we obtain the next-step Lagrange multipliers (3.42)

$$\theta^{view(2)} = 5.52, \tag{3.49}$$

and then the scenario-probability expectation and covariance reads

$$\boldsymbol{\mu}_{\boldsymbol{X}}^{(2)} \equiv \left(\begin{smallmatrix} 0.66 \\ -0.39 \end{smallmatrix}\right), \quad \boldsymbol{\sigma}_{\boldsymbol{X}}^{2(2)} \equiv \left(\begin{smallmatrix} 0.18 & 0.11 \\ 0.11 & 0.23 \end{smallmatrix}\right). \tag{3.50}$$

Note how the new estimate $\hat{\theta}^{view(2)}$ is closer to the true parameter $\bar{\theta}^{view}$ than the older one $\hat{\theta}^{view(1)} \equiv \hat{\theta}^{view}$

$$|\theta^{view*} - \theta^{view(2)}| = 0.1 < 0.6 = |\theta^{view*} - \theta^{view(1)}|. \tag{3.51}$$

Indeed the whole new scenario-probability distribution is much more similar to true MRE updated distribution than its older counterpart, compare Figure 3.2 with Figure 3.1.

3.4.2 Equivalent stopping criteria

According to the above discussion, an indicator of convergence for the iterative approach are the next-step Lagrange multipliers $\boldsymbol{\epsilon}^{view(i+1)}$: the more the Euclidean norm (3.47) is close to zero the more the new i-th base $f^{(i)}_{\boldsymbol{X}}$ (3.44) is close to the true MRE updated distribution $\bar{f}_{\boldsymbol{X}}$. In particular, it turns out that the number of iterations needed to reach a given confidence level of accuracy of $\bar{f}_{\boldsymbol{X}}$ depends on how much the views (3.2), as quantified by the features $\boldsymbol{\eta}^{view}$, are far from being satisfied by the new base i-th $f^{(i)}_{\boldsymbol{X}}$.

This intuition is enforced by computing the **views intensity** [Meucci, 2019], which is defined, for a given base distribution $\underline{f}_{\boldsymbol{X}}$ (3.1), as the Euclidean norm of gradient of the minimal relative entropy $\mathcal{E}(\bar{f}_{\boldsymbol{X}}||\underline{f}_{\boldsymbol{X}})$ (17) with respect to the features $\boldsymbol{\eta}^{view}$ and which explicitly reads [A.3.5]

$$Intensity(\boldsymbol{\eta}^{view}, \underline{f}_{\boldsymbol{X}}) \equiv ||\nabla_{\boldsymbol{\eta}^{view}}\mathcal{E}(\bar{f}_{\boldsymbol{X}}||\underline{f}_{\boldsymbol{X}})|| = ||\boldsymbol{\theta}^{view}||. \tag{3.52}$$

In particular, if we consider $f^{(i)}_{\boldsymbol{X}}$ (3.44) as base distribution, the views intensity is fully identified by the norm of the next-step Lagrange multipliers $||\boldsymbol{\epsilon}^{view(i+1)}||$. This is not surprising, indeed if the new base $\bar{f}^{(i)}_{\boldsymbol{X}}$ satisfied the views, then the updated $\bar{f}_{\boldsymbol{X}}$ should coincide with the base distribution itself, or $\boldsymbol{\epsilon}^{view(i+1)} = \mathbf{0}$ (26). This would imply that the views intensity would be null in turn, or $Intensity(\boldsymbol{\eta}^{view}, f^{(i)}_{\boldsymbol{X}}) = 0$.

Example 3.4. We continue from Example 3.3. The views intensity from going to the original base $f^{(0)}_{\boldsymbol{X}} \equiv \underline{f}_{\boldsymbol{X}}$ to the updated distribution

$$Intensity(\boldsymbol{\eta}^{view}, \underline{f}_{\boldsymbol{X}}) \approx |\theta^{view(1)}| = 5.59 \tag{3.53}$$

is higher than the views intensity from going to the new base $f^{(1)}_{\boldsymbol{X}}$ to the updated distribution

$$Intensity(\boldsymbol{\eta}^{view}, f^{(1)}_{\boldsymbol{X}}) \approx |\theta^{view(2)} - \theta^{view(1)}| = 0.02. \tag{3.54}$$

Another indicator of convergence for the iterative approach is the statistical significance of the updated probabilities as in (3.32)

$$\boldsymbol{p}^{(i+1)} \propto \underline{\boldsymbol{p}} \circ e^{\boldsymbol{\epsilon}^{view(i+1)\prime}\zeta^{view}(\mathcal{X}^{(i)})}, \tag{3.55}$$

where $\mathcal{X}^{(i)}$ denote the i-th i.i.d. HMC scenarios stemming from the new i-th base $f^{(i)}_{\boldsymbol{X}}$ (3.44) as in (3.19), or

$$\mathcal{X}^{(i)} \equiv hmc\_sampler(g^{(i)}_{\boldsymbol{X}}, \bar{\jmath}); \tag{3.56}$$

and $\underline{\boldsymbol{p}}$ are the uniform probabilities (3.20).

As a matter of fact, the intuition suggests that the update of the probabilities from $\underline{\boldsymbol{p}}$ should be mild once $f^{(i)}_{\boldsymbol{X}}$ is enough close to $\bar{f}_{\boldsymbol{X}}$.

This intuition is enforced by computing the **relative effective number of scenarios** (ENS) [Meucci, 2012a] which is an index between 0 and 1 defined as the exponential of the discrete Shannon entropy over the number of scenarios $\bar{j}$

$$ENS(\boldsymbol{p}) \equiv \frac{e^{-\boldsymbol{p}' \ln \boldsymbol{p}}}{\bar{j}}, \tag{3.57}$$

and quantifying how far the probabilities are from being uniform.

It turns out that the ENS of the updated probabilities (3.32) explicitly reads [A.3.6]

$$ENS(\boldsymbol{p}^{(i+1)}) \approx e^{-\mathcal{E}(f_{\boldsymbol{X}}^{(i+1)} \| f_{\boldsymbol{X}}^{(i)})}. \tag{3.58}$$

Hence, the relative effective number of scenarios is fully identified by the relative entropy $\mathcal{E}(f_{\boldsymbol{X}}^{(i+1)} \| f_{\boldsymbol{X}}^{(i)})$. This is not surprising, as a matter of fact, if $\mathcal{E}(f_{\boldsymbol{X}}^{(i+1)} \| f_{\boldsymbol{X}}^{(i)})$ were null, then the $(i+1)$-th update $f_{\boldsymbol{X}}^{(i+1)}$ would be the same as the i-th $f_{\boldsymbol{X}}^{(i)}$ in turn, or $\boldsymbol{\epsilon}^{view(i+1)} = \mathbf{0}$ (26). This would imply that the relative effective number of scenarios would be one, or $\boldsymbol{p}^{(i+1)} = \underline{\boldsymbol{p}}$.

From (3.58) it is easy to verify empirically that two indicators we introduced are equivalent: the lower the views intensity (3.52) and the higher the effective number of scenarios (3.58)

$$||\boldsymbol{\epsilon}^{view(i+1)}|| \approx 0 \quad \Leftrightarrow \quad ENS(\boldsymbol{p}^{(i+1)}) \approx 1. \tag{3.59}$$

Example 3.5. We continue from Example 3.4. The effective number of scenarios (3.58) of the updated probabilities $\boldsymbol{p}^{(1)} \equiv \bar{\boldsymbol{p}}$ (3.32) stemming from the original HMC base scenarios $\mathcal{X}^{(0)} \equiv \underline{\mathcal{X}}$

$$ENS(\boldsymbol{p}^{(1)}) = 5.49\%, \tag{3.60}$$

is lower than the effective number of scenarios of the updated probabilities $\boldsymbol{p}^{(2)}$ (3.55) stemming from the new HMC scenarios $\mathcal{X}^{(1)}$

$$ENS(\boldsymbol{p}^{(2)}) = 99.99\%. \tag{3.61}$$

To conclude, we summarize the steps from (3.11) to (3.13) of the iterative MRE in the following table.

$(\{\overline{\mathcal{X}}, \bar{\boldsymbol{p}}\}, \hat{\boldsymbol{\theta}}^{view}) = Iterative.MRE(\zeta^{view}, \boldsymbol{\eta}^{view}, \underline{g}_{\boldsymbol{X}}, \bar{j})$	
0. Initialize	$\hat{\boldsymbol{\theta}}^{view} \leftarrow \mathbf{0}_{\bar{k}\times 1}$
1. Update scenarios	$\{\overline{\mathcal{X}}, \underline{\boldsymbol{p}}\} \overset{\text{HMC}}{\Leftarrow} (\underline{g}_{\boldsymbol{X}}(\cdot)e^{\hat{\boldsymbol{\theta}}^{view\prime}\zeta^{view}(\cdot)}, \bar{j})$ (3.56)
2. Update Lagr. mult.	$\begin{cases} \hat{\boldsymbol{\epsilon}}^{view} \Leftarrow (\zeta^{view}, \boldsymbol{\eta}^{view}, \{\overline{\mathcal{X}}, \underline{\boldsymbol{p}}\}) \text{ (3.23)} \\ \hat{\boldsymbol{\theta}}^{view} \leftarrow \hat{\boldsymbol{\theta}}^{view} + \hat{\boldsymbol{\epsilon}}^{view} \text{ (3.42)} \end{cases}$
3. Update prob.	$\bar{\boldsymbol{p}} \Leftarrow (\hat{\boldsymbol{\epsilon}}^{view}, \zeta^{view}, \{\overline{\mathcal{X}}, \underline{\boldsymbol{p}}\})$ (3.55)
4. If convergence, output $(\{\overline{\mathcal{X}}, \bar{\boldsymbol{p}}\}, \hat{\boldsymbol{\theta}}^{view})$; else and go to 1	

Table 3.2: Iterative MRE algorithm

Convergence in the above routine occurs when the Euclidean norm $||\hat{\boldsymbol{\epsilon}}^{view}||$, is smaller than a required threshold $0 < \delta \ll 1$, or equivalently (3.59), the relative effective number of scenarios $ENS(\bar{\boldsymbol{p}})$ is higher than $1 - \delta$. Note in particular note that, according to the non-parametric counterpart (3.34), by construction the views (3.2) are satisfied in sample by the scenario-probability distribution (3.6) identified by the outcome $\{\overline{\mathcal{X}}, \bar{\boldsymbol{p}}\}$.

3.5 Comparison

In order to highlight the benefits of the iterative MRE (3.2) we consider the respective i-th estimators of the Lagrange multipliers, which are a function of the randomized base sample (3.19) representing the data generating process (DGP)

$$\boldsymbol{\Theta}^{view(i)} = d^{(i)}(\{\boldsymbol{X}^{(j)}\}_{j=1}^{\bar{j}}). \tag{3.62}$$

In particular, we can summarize the accuracy and dispersion of the estimators through their bias

$$bias^{(i)} \equiv ||\mathbb{E}\{\boldsymbol{\Theta}^{view(i)}\} - \boldsymbol{\theta}^{view}||, \tag{3.63}$$

and inefficiency

$$inef^{(i)} \equiv \sqrt{\mathbb{E}\{||\boldsymbol{\Theta}^{view(i)} - \mathbb{E}\{\boldsymbol{\Theta}^{view(i)}\}||^2\}}. \tag{3.64}$$

Moreover, we can also consider the distribution of the loss implied by the estimator

$$Loss^{(i)} \equiv ||\boldsymbol{\Theta}^{view(i)} - \boldsymbol{\theta}^{view}||^2, \tag{3.65}$$

and we evaluate the goodness of ours estimators, through its expectation, or error $er \equiv \mathbb{E}^f\{Loss\}$, which is connected to the bias and inefficiency according to the following relationship

$$er^{(i)} = inef^{(i)2} + bias^{(i)2}. \tag{3.66}$$

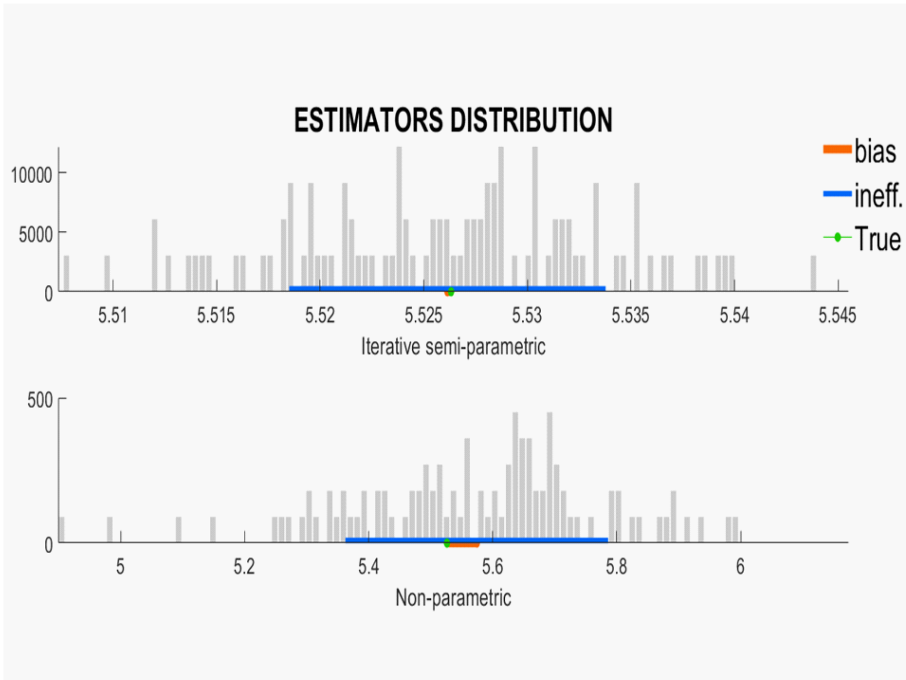

Figure 3.3: Non-parametric vs iterative semi-parametric: estimator distributions for the Lagrange multipliers

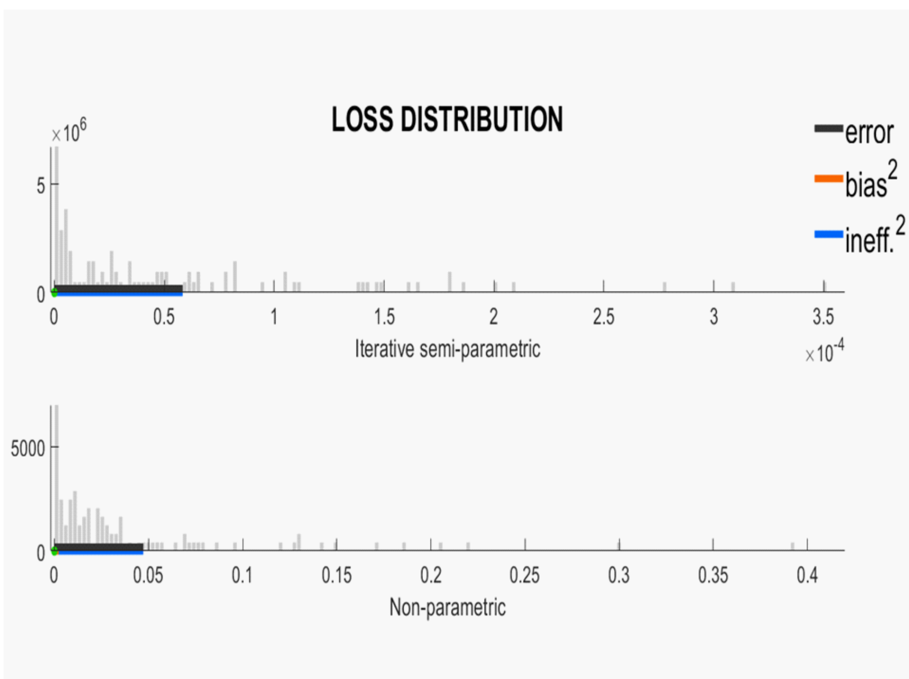

Figure 3.4: Non-parametric vs iterative semi-parametric: loss distributions for the Lagrange multipliers

It turns out empirically that at each iteration the distribution of the estimators, or equivalently the loss, becomes closer to the true Lagrange multipliers, or equivalently to zero, hence yielding a better approximation than the non-parametric approach (3.23).

Example 3.6. We continue from Example 3.3. We resort to simulations and compute the error (3.66), bias (3.63) and inefficiency (3.64) corresponding to the estimators of both non-parametric and iterative semi-parametric approach. As highlighted in the following table, the iterative semi-parametric approach provide estimates with lower bias and inefficiency (and hence a lower error) than the non-parametric counterpart.

	Non-parametric MRE	Iterative MRE
bias	4.81×10^{-2}	1.78×10^{-4}
inef	2.12×10^{-1}	7.64×10^{-3}
err	4.72×10^{-2}	5.83×10^{-5}

Table 3.3: Non-parametric vs iterative MRE: bias, inefficiency and error

See also Figures 3.3 and 3.4.

3.6 Conclusions

We investigated how minimum relative entropy models can be estimated in a non-parametric setting within the exponential family class. The main insight is the numerical approximation of the updated solution via Hamiltonian Monte Carlo simulations that can be iterated in order to reach a good approximation of the parameters that drive the MRE updated distribution, i.e. the Lagrange multipliers. Another theoretical contribution is the equivalence relationship between of the statistical significance of the non-parametric framework with respect to the intensity of the views.

Finally within the normal framework introduced in Chapter 1, we showed how the iterative implementation significantly reduces the estimation error of the estimators of the true unknown Lagrange multipliers, yielding a more precise approximation than the non-parametric approach by [Meucci, 2008].

Bibliography

[Aitchison and Dunsmore, 1975] Aitchison, J. and Dunsmore, I. R. (1975). *Statistical Prediction Analysis*. Cambridge University Press.

[Almgren and Chriss, 2006] Almgren, R. and Chriss, N. (2006). Optimal portfolios from ordering information. *Journal of Risk*, 9:1–47.

[Amari and Nagaoka, 2000] Amari, S. and Nagaoka, H. (2000). *Methods of Information Geometry*. American Mathematical Society.

[Amari, 2016] Amari, S.-i. (2016). *Information Geometry and Its Applications*, volume 194. Springer.

[Anderson, 1984] Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*. Wiley, 2nd edition.

[Asness et al., 2013a] Asness, C., Frazzini, A., and Pedersen, L. (2013a). Quality minus junk. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2312432. Working Paper.

[Asness et al., 2013b] Asness, C. S., Moskowitz, T. J., and Pedersen, L. H. (2013b). Value and momentum everywhere. *Journal of Finance*, 68(3):929–985.

[Avellaneda, 1999] Avellaneda, M. (1999). Minimum-entropy calibration of asset-pricing models. *International Journal of Theoretical and Applied Finance*, 1:447–472.

[Ben-Tal and Nemirovski, 2001] Ben-Tal, A. and Nemirovski, A. (2001). *Lectures on modern convex optimization: analysis, algorithms, and engineering applications*. Society for Industrial and Applied Mathematics.

[Berg, 2004] Berg, B. A. (2004). *Markov chain Monte Carlo simulations and their statistical analysis: with web-based Fortran code*. World Scientific Publishing Company.

[Black and Litterman, 1990] Black, F. and Litterman, R. (1990). Asset allocation: combining investor views with market equilibrium. *Goldman Sachs Fixed Income Research*.

[Black and Scholes, 1973] Black, F. and Scholes, M. S. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, 81:637–654.

[Breuer and Csiszar, 2013] Breuer, T. and Csiszar, I. (2013). Systematic stress tests with entropic plausibility constraints. *Journal of banking and finance*, 37:1552–1559.

[Caticha and Giffin, 2006] Caticha, A. and Giffin, A. (2006). Updating probabilities. *Bayesian Inference and Maximum Entropy Methods in Science and Engineering, AIP Conf. Proc.*

[Chao et al., 2015] Chao, W.-L., Solomon, J., Michels, D., and Sha, F. (2015). Exponential integration for hamiltonian monte carlo. In *Proceedings of the 32nd International Conference on Machine Learning (ICML-15)*, pages 1142–1151.

[Chib and Greenberg, 1995] Chib, S. and Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistician*, 49:327–335.

[Cont, 2007] Cont, R. Ans Tankov, P. (2007). Recovering exponential Lévy models from option prices: Regularization of an ill-posed inverse problem. *SIAM Journal on Control and Optimization*, 45:1–25.

[Cover and Thomas, 2006] Cover, T. M. and Thomas, J. A. (2006). *Elements of Information Theory*. Wiley, 2nd edition.

[Coxeter, 1998] Coxeter, H. S. M. (1998). *Non-euclidean geometry*. Cambridge University Press.

[D'Amico et al., 2003] D'Amico, M., Fusai, G., and Tagliani, A. (2003). Valuation of exotic options using moments. *Operational Research*, 2:157–186.

[DeMiguel et al., 2009] DeMiguel, V., Garlappi, L., and Uppal, R. (2009). How inefficient is the 1/n asset-allocation strategy? *Working Paper*.

[Friedman et al., 2012] Friedman, C., Zhang, Y., and Cao, W. (2012). Fat tails via utility based entropy. *Risk*, 25(10):68.

[Geweke, 1999] Geweke, J. (1999). Using simulation methods for Bayesian econometric models: Inference, development and communication. *Econometric Reviews*, 18:1–126.

[Glasserman and Xu, 2013] Glasserman, P. and Xu, X. (2013). Robust portfolio control with stochastic factor dynamics. *Operations Research*, 61(4):874–893.

[Grinold and Kahn, 1999] Grinold, R. C. and Kahn, R. (1999). *Active Portfolio Management. A Quantitative Approach for Producing Superior Returns and Controlling Risk*. McGraw-Hill, 2nd edition.

[Hansen and Sargent, 2007] Hansen, L. P. and Sargent, T. J. (2007). *Robustness*. Princeton University Press.

[Jaakkola, 1999] Jaakkola, T. (1999). Maximum entropy estimation. http://people.csail.mit.edu/tommi/papers.html. Machine learning seminar notes.

[Jagannathan and Ma, 2003] Jagannathan, R. and Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *Journal of Finance*, 58:1651–1683.

[Jaynes, 1957a] Jaynes, E. T. (1957a). Information theory and statistical mechanics. *Physical review*, 106(4):620.

[Jaynes, 1957b] Jaynes, E. T. (1957b). Information theory and statistical mechanics. ii. *Physical review*, 108(2):171.

[Jobson and Korkie, 1981] Jobson, J. D. and Korkie, B. M. (1981). Performance hypothesis testing with the Sharpe and Treynor measures. *Journal of Finance*, 36(4):889–908.

[Kan and Zhou, 2006] Kan, R. and Zhou, G. (2006). Optimal estimation for economic gains: Portfolio choice with parameter uncertainty. *Journal of Financial and Quantitative Analysis*.

[Kleywegt and Shapiro, 2001] Kleywegt, A. J. and Shapiro, A. (2001). Stochastic optimization. *Handbook of Industrial Engineering*, pages 2625–2649.

[Kriz and Talacko, 1968] Kriz, T. A. and Talacko, J. V. (1968). Equivalence of the maximum likelihood estimator to a minimum entropy estimator. *Trabajos de estadística y de investigación operativa*, 19(1-2):55–65.

[Lehmann and Casella, 1998] Lehmann, E. L. and Casella, G. (1998). *Theory of Point Estimation*. Springer, 2nd edition.

[Lobo et al., 2007] Lobo, M., Fazel, M., and Boyd, B. (2007). Portfolio optimization with linear and fixed transaction costs. *Annals of Operations Research*, 152(1):341–365.

[Magnus and Neudecker, 1979] Magnus, J. R. and Neudecker, H. (1979). The commutation matrix: Some properties and applications. *Annals of Statistics*, 7:381–394.

[Malouf, 2002] Malouf, R. (2002). A comparison of algorithms for maximum entropy parameter estimation. In *proceedings of the 6th conference on Natural language learning-Volume 20*, pages 1–7. Association for Computational Linguistics.

[Markowitz, 1952] Markowitz, H. M. (1952). Portfolio selection. *Journal of Finance*, 7(1):77–91.

[Menchero et al., 2013] Menchero, J., Morozov, A., and Pasqua, A. (2013). Predicting risk at short horizons. http://www.lqg.org.uk/wp-content/uploads/2013/02/Predicting-Risk-at-Short-Horizons-Model-Insight.pdf.

[Meucci, 2005] Meucci, A. (2005). Risk and asset allocation. http://arpm.co. Springer Finance.

[Meucci, 2008] Meucci, A. (2008). Fully Flexible Views: Theory and practice. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1213325. Risk, 21(10), 97-102.

[Meucci, 2010] Meucci, A. (2010). The Black-Litterman approach: Original model and extensions. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1117574. The Encyclopedia of Quantitative Finance, Wiley.

[Meucci, 2011] Meucci, A. (2011). Mixing probabilities and kernels via Entropy Pooling. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1944303. Garp Risk Professional, 32-36.

[Meucci, 2012a] Meucci, A. (2012a). Effective number of scenarios with Fully Flexible Probabilities. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1971808. GARP Risk Professional, 45-46.

[Meucci, 2012b] Meucci, A. (2012b). Stress-testing with fully flexible causal inputs. Risk, 25(4), 61-65.

[Meucci, 2013] Meucci, A. (2013). Estimation and stress-testing via time- and market-conditional flexible probabilities. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2312126. Working Paper.

[Meucci, 2019] Meucci, A. (2019). ARPM Lab. https://www.arpm.co/lab.

[Meucci et al., 2014] Meucci, A., Ardia, D., and Colasante, M. (2014). Portfolio construction and systematic trading with Factor Entropy Pooling. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1742559. Risk, 27 (5), 56-61.

[Meucci et al., 2011] Meucci, A., Ardia, D., and Keel, S. (2011). Fully flexible views in multivariate normal markets.

[Meucci and Nicolosi, 2016] Meucci, A. and Nicolosi, M. (2016). Dynamic portfolio management with views at multiple horizons. *Applied Mathematics and Computation*, 274:495–518.

[Mina and Xiao, 2001] Mina, J. and Xiao, J. (2001). Return to RiskMetrics: The evolution of a standard. *RiskMetrics publications*.

[Moskowitz et al., 2012] Moskowitz, T., Ooi, Y. H., and Pedersen, L. H. (2012). Time series momentum. *Journal of Financial Economics*, 104(228-250):2.

[Murphy, 2007] Murphy, K. P. (2007). Conjugate Bayesian analysis of the gaussian distribution. *def*, 1(2σ2):16.

[Neal et al., 2011] Neal, R. M. et al. (2011). Mcmc using hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo*, 2(11).

[Park, 2010] Park, S. C. (2010). The moving average ratio and momentum. *Financial Review*, 45(2):415–447.

[Pezier, 2007] Pezier, J. (2007). Global portfolio optimization revisited: A least discrimination alternantive to Black-Litterman. *ICMA Centre Discussion Papers in Finance*.

[Qian and Gorman, 2001] Qian, E. and Gorman, S. (2001). Conditional distribution in portfolio theory. *Financial Analyst Journal*, 57:44–51.

[RiskMetrics, 1996] RiskMetrics (1996). *RiskMetricsTechnical Document*. J.P.Morgan and Reuters Publications, 4th edition.

[Schofield, 2007] Schofield, E. (2007). *Fitting maximum-entropy models on large sample spaces*. PhD thesis, Graz University of Technology Austria.

[Stein, 1955] Stein, C. (1955). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. *Proceedings of the 3rd Berkeley Symposium on Probability and Statistics*.

[Wang and Kochard, 2011] Wang, P. and Kochard, L. (2011). Using a z-score approach to combine value and momentum in tactical asset allocation. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1726443. Working Paper.

Appendix 1

Here we discuss some technical results of Chapter 1.

A.1.1 Exponential family distributions: MRE update under exponential-family base

Let us suppose the case where the base distribution (1) belongs to an exponential family class as in (23)

$$\underline{f}_{\boldsymbol{X}} \quad \Leftrightarrow \quad \mathit{Exp}(\underline{\boldsymbol{\theta}}_{\boldsymbol{X}}, \tau, h), \tag{A.1.1}$$

for some vector $\underline{\boldsymbol{\theta}}_{\boldsymbol{X}} \equiv (\underline{\theta}_{\boldsymbol{X};1}, \ldots, \underline{\theta}_{\boldsymbol{X};\bar{l}})' \in \mathbb{R}^{\bar{l}}$ of base canonical parameters**;** sufficient statistics $\tau(\boldsymbol{x}) \equiv (\tau_1(\boldsymbol{x}), \ldots, \tau_{\bar{l}}(\boldsymbol{x}))'$ and reference measure $h(\boldsymbol{x}) > 0$.

Then, under views on generalized expectations (22), the updated distribution $\bar{f}_{\boldsymbol{X}}$ (16) is an exponential family distribution (23) with respect to the base $\underline{f}_{\boldsymbol{X}}$ as reference measure.

In particular, since the base distribution belongs to the exponential family in turn (A.1.1), the updated distribution $\bar{f}_{\boldsymbol{X}}$ is also an exponential family distribution with respect to the reference measure h.

Indeed it is easy to verify that the updated distribution $\bar{f}_{\boldsymbol{X}}$ (24) can be written as follows

$$\begin{aligned} \bar{f}_{\boldsymbol{X}}(\boldsymbol{x}) &= \underline{f}_{\boldsymbol{X}}(\boldsymbol{x}) e^{\boldsymbol{\theta}^{view\prime}\zeta^{view}(\boldsymbol{x}) - \psi_{\underline{f}_{\boldsymbol{X}},\zeta^{view}}(\boldsymbol{\theta}^{view})} \\ &= h(\boldsymbol{x}) \exp(\underline{\boldsymbol{\theta}}'_{\boldsymbol{X}}\tau(\boldsymbol{x}) - \psi_{h,\tau}(\underline{\boldsymbol{\theta}}_{\boldsymbol{X}})) \times \exp(\boldsymbol{\theta}^{view\prime}\zeta^{view}(\boldsymbol{x}) - \psi_{\underline{f}_{\boldsymbol{X}},\zeta^{view}}(\boldsymbol{\theta}^{view})) \\ &= h(\boldsymbol{x}) \exp(\bar{\boldsymbol{\theta}}'_{\boldsymbol{X}}\bar{\tau}(\boldsymbol{x}) - \psi_{h,\bar{\tau}}(\bar{\boldsymbol{\theta}}_{\boldsymbol{X}})), \end{aligned} \tag{A.1.2}$$

where we defined the new canonical parameters $\bar{\boldsymbol{\theta}}_{\boldsymbol{X}}$ and sufficient statistics $\bar{\tau}$ via juxtaposition

$$\bar{\boldsymbol{\theta}}_{\boldsymbol{X}} \equiv \begin{pmatrix} \underline{\boldsymbol{\theta}}_{\boldsymbol{X}} \\ \boldsymbol{\theta}^{view} \end{pmatrix}, \quad \bar{\tau}(\boldsymbol{x}) \equiv \begin{pmatrix} \tau(\boldsymbol{x}) \\ \zeta^{view}(\boldsymbol{x}) \end{pmatrix}; \tag{A.1.3}$$

and where we defined $\psi_{h,\bar{\tau}}$ as the sum of log-partition functions

$$\psi_{h,\bar{\tau}}(\bar{\boldsymbol{\theta}}_{\boldsymbol{X}}) \equiv \psi_{h,\tau}(\underline{\boldsymbol{\theta}}_{\boldsymbol{X}}) + \psi_{\underline{f}_{\boldsymbol{X}},\zeta^{view}}(\boldsymbol{\theta}^{view}). \tag{A.1.4}$$

This means that the updated distribution $\bar{f}_{\boldsymbol{X}}$ is also an exponential family distribution of the form

$$\bar{f}_{\boldsymbol{X}} \quad \Leftrightarrow \quad \mathit{Exp}(\boldsymbol{\theta}^{view}, \zeta^{view}, \underline{f}_{\boldsymbol{X}}) \quad \Leftrightarrow \quad \mathit{Exp}(\bar{\boldsymbol{\theta}}_{\boldsymbol{X}}, \bar{\tau}, h). \tag{A.1.5}$$

As a matter of fact note that $\psi_{h,\bar{\tau}}$ (A.1.4) is a log-partition function in turn, since we have

$$\begin{aligned}\psi_{h,\bar{\tau}}(\bar{\boldsymbol{\theta}}_{\boldsymbol{X}}) &= \psi_{h,\tau}(\underline{\boldsymbol{\theta}}_{\boldsymbol{X}}) + \ln \int_{\mathbb{R}^{\bar{n}}} \underline{f}_{\boldsymbol{X}}(\boldsymbol{x})\, e^{\boldsymbol{\theta}^{view\prime}\zeta^{view}(\boldsymbol{x})} d\boldsymbol{x} \\ &= \psi_{h,\tau}(\underline{\boldsymbol{\theta}}_{\boldsymbol{X}}) + \ln \int_{\mathbb{R}^{\bar{n}}} h(\boldsymbol{x})\, e^{\underline{\boldsymbol{\theta}}'_{\boldsymbol{X}}\tau(\boldsymbol{x}) - \psi_{h,\tau}(\underline{\boldsymbol{\theta}}_{\boldsymbol{X}})} e^{\boldsymbol{\theta}^{view\prime}\zeta^{view}(\boldsymbol{x})} d\boldsymbol{x} \\ &= \psi_{h,\tau}(\underline{\boldsymbol{\theta}}_{\boldsymbol{X}}) - \psi_{h,\tau}(\underline{\boldsymbol{\theta}}_{\boldsymbol{X}}) + \ln \int_{\mathbb{R}^{\bar{n}}} h(\boldsymbol{x})\, e^{\bar{\boldsymbol{\theta}}'_{\boldsymbol{X}}\bar{\tau}(\boldsymbol{x})} d\boldsymbol{x} \\ &= \ln \int_{\mathbb{R}^{\bar{n}}} h(\boldsymbol{x})\, e^{\bar{\boldsymbol{\theta}}'_{\boldsymbol{X}}\bar{\tau}(\boldsymbol{x})} d\boldsymbol{x}. \end{aligned} \quad (A.1.6)$$

This means that in general the updated distribution $\bar{f}_{\boldsymbol{X}}$ does *not* belong necessarily to the same exponential family class of the base counterpart $\underline{f}_{\boldsymbol{X}}$ (A.1.1).

However, *if* the view functions ζ^{view} can be expressed as a linear combination of the sufficient statistics τ, or

$$\zeta^{view}(\boldsymbol{x}) = \boldsymbol{\zeta}\tau(\boldsymbol{x}), \quad (A.1.7)$$

for some suitable $\bar{k} \times \bar{l}$ matrix $\boldsymbol{\zeta}$, then we can redefine the new canonical parameters $\bar{\boldsymbol{\theta}}_{\boldsymbol{X}}$ and sufficient statistics $\bar{\tau}$ as follows

$$\bar{\boldsymbol{\theta}}_{\boldsymbol{X}} \equiv \underline{\boldsymbol{\theta}}_{\boldsymbol{X}} + \boldsymbol{\zeta}'\boldsymbol{\theta}^{view}, \quad \bar{\tau}(\boldsymbol{x}) \equiv \tau(\boldsymbol{x}), \quad (A.1.8)$$

which means that the updated distribution $\bar{f}_{\boldsymbol{X}}$ and the base counterpart $\underline{f}_{\boldsymbol{X}}$ are *conjugate*, i.e. $\bar{f}_{\boldsymbol{X}}$ belongs to the *same* exponential family class of $\underline{f}_{\boldsymbol{X}}$

$$\bar{f}_{\boldsymbol{X}} \quad \Leftrightarrow \quad Exp(\boldsymbol{\theta}^{view}, \zeta^{view}, \underline{f}_{\boldsymbol{X}}) \quad \Leftrightarrow \quad Exp(\underline{\boldsymbol{\theta}}_{\boldsymbol{X}} + \boldsymbol{\zeta}\boldsymbol{\theta}^{view}, \tau, h). \quad (A.1.9)$$

This also means from (A.1.4)

$$\psi_{h,\tau}(\bar{\boldsymbol{\theta}}_{\boldsymbol{X}}) = \psi_{h,\tau}(\underline{\boldsymbol{\theta}}_{\boldsymbol{X}}) + \psi_{\underline{f}_{\boldsymbol{X}},\zeta^{view}}(\boldsymbol{\theta}^{view}). \quad (A.1.10)$$

Note how this result generalizes the normal case (1.15)-(1.5).

A.1.2 Normal MRE update: canonical representation

The pdf of the normal base distribution (1.10) can be rewritten in canonical form as in (24)

$$\underline{f}_{\boldsymbol{X}}(\boldsymbol{x}) = (2\pi)^{-\frac{\bar{n}}{2}} \exp(\underline{\boldsymbol{\theta}}'_{\boldsymbol{X};\mu}\boldsymbol{x} + vec(\underline{\boldsymbol{\theta}}_{\boldsymbol{X};\sigma,\sigma})'vec(\boldsymbol{x}\boldsymbol{x}') - \psi^N(\underline{\boldsymbol{\theta}}_{\boldsymbol{X}})), \quad (A.1.11)$$

where $\underline{\boldsymbol{\theta}}_{\boldsymbol{X};\mu}$ and $\underline{\boldsymbol{\theta}}_{\boldsymbol{X};\sigma,\sigma}$ are the base canonical coordinates (1.11) and where ψ^N denotes the log-partition function as in (25) with respect to the reference measure $h(\boldsymbol{x}) \equiv (2\pi)^{-\bar{n}/2}$ and statistics τ (1.12), which explicitly reads

$$\begin{aligned}\psi^N(\underline{\boldsymbol{\theta}}_{\boldsymbol{X}}) &\equiv \psi_{h,\tau}(\underline{\boldsymbol{\theta}}_{\boldsymbol{X}}) \equiv \ln \int_{\mathbb{R}^{\bar{n}}} e^{\underline{\boldsymbol{\theta}}'_{\boldsymbol{X}}\tau(\boldsymbol{x})} h(\boldsymbol{x})\, d\boldsymbol{x} \\ &= -\tfrac{1}{4}\underline{\boldsymbol{\theta}}'_{\boldsymbol{X};\mu}(\underline{\boldsymbol{\theta}}_{\boldsymbol{X};\sigma,\sigma})^{-1}\underline{\boldsymbol{\theta}}_{\boldsymbol{X};\mu} - \tfrac{1}{2}\ln\det(-2\underline{\boldsymbol{\theta}}_{\boldsymbol{X};\sigma,\sigma}),\end{aligned} \quad (A.1.12)$$

see also [Amari and Nagaoka, 2000] and [Amari, 2016].

Now, since the moments conditions (1.9) are in the expectation format (22), the ensuing updated distribution is an exponentially-twisted normal (23), and must read

$$\begin{aligned}\bar{f}_{\boldsymbol{X}}(\boldsymbol{x}) &= \underline{f}_{\boldsymbol{X}}(\boldsymbol{x})e^{\boldsymbol{\theta}^{view\prime}\zeta^{view}(\boldsymbol{x})-\psi(\boldsymbol{\theta}^{view})} \\ &= \underline{f}_{\boldsymbol{X}}(\boldsymbol{x})\exp(\boldsymbol{\theta}_{\mu}^{view\prime}\boldsymbol{\zeta}_{\mu}\boldsymbol{x} + vec(\boldsymbol{\theta}_{\sigma}^{view})'(\boldsymbol{\zeta}_{\sigma}\otimes\boldsymbol{\zeta}_{\sigma})vec(\boldsymbol{x}\boldsymbol{x}') - \psi(\boldsymbol{\theta}^{view})) \\ &= (2\pi)^{-\frac{\bar{n}}{2}}\exp(\bar{\boldsymbol{\theta}}_{\boldsymbol{X};\mu}'\boldsymbol{x} + vec(\bar{\boldsymbol{\theta}}_{\boldsymbol{X};\sigma})'vec(\boldsymbol{x}\boldsymbol{x}') - (\psi(\boldsymbol{\theta}^{view}) + \psi^{N}(\underline{\boldsymbol{\theta}}_{\boldsymbol{X}}))),\end{aligned} \tag{A.1.13}$$

as follows by replacing the normal base $\underline{f}_{\boldsymbol{X}}$ (A.1.11) and using the definition of updated canonical coordinates $\bar{\boldsymbol{\theta}}_{\boldsymbol{X}}$ (1.5).

Hence, according to the canonical representation of normal distributions as in (A.1.11), the updated $\bar{f}_{\boldsymbol{X}}$ must be normal as in (1.15).

Moreover, because of the uniqueness of the canonical representation, the log-partition function of the updated distribution as in (25) must satisfy the following condition

$$\psi^{N}(\bar{\boldsymbol{\theta}}_{\boldsymbol{X}}) = \psi(\boldsymbol{\theta}^{view}) + \psi^{N}(\underline{\boldsymbol{\theta}}_{\boldsymbol{X}}). \tag{A.1.14}$$

See more details in [A.1.3] and generalizations in [A.1.1].

A.1.3 Normal MRE update: dual Lagrangian, location and dispersion

From the canonical representation of the normal updated distribution (A.1.13), we obtain the dual Lagrangian (28) corresponding to the views (1.9)

$$\begin{aligned}\mathcal{L}(\boldsymbol{t};\boldsymbol{\eta}^{view}) &\equiv \psi(\boldsymbol{t}) - \boldsymbol{t}_{\mu}'\boldsymbol{\eta}_{\mu}^{view} - \mathrm{tr}(\boldsymbol{t}_{\sigma,\sigma}'\boldsymbol{\eta}_{\sigma,\sigma}^{view}) \\ &= \psi^{N}(\begin{pmatrix}\underline{\boldsymbol{\theta}}_{\boldsymbol{X};\mu} + \boldsymbol{\zeta}_{\mu}'\boldsymbol{t}_{\mu} \\ vec(\underline{\boldsymbol{\theta}}_{\boldsymbol{X};\sigma,\sigma} + \boldsymbol{\zeta}_{\sigma}'\boldsymbol{t}_{\sigma,\sigma}\boldsymbol{\zeta}_{\sigma})\end{pmatrix}) - \psi^{N}(\underline{\boldsymbol{\theta}}_{\boldsymbol{X}}) - \boldsymbol{t}_{\mu}'\boldsymbol{\eta}_{\mu}^{view} - \mathrm{tr}(\boldsymbol{t}_{\sigma,\sigma}'\boldsymbol{\eta}_{\sigma,\sigma}^{view}),\end{aligned} \tag{A.1.15}$$

where ψ^{N} denotes the canonical normal log-partition function (A.1.12).

Now, using the relationship between canonical coordinates and expectation (1.17)

$$\boldsymbol{\mu}_{\boldsymbol{X}} = -\frac{1}{2}(\underline{\boldsymbol{\theta}}_{\boldsymbol{X};\sigma,\sigma} + \boldsymbol{\zeta}_{\sigma}'\boldsymbol{t}_{\sigma,\sigma}\boldsymbol{\zeta}_{\sigma})^{-1}(\underline{\boldsymbol{\theta}}_{\boldsymbol{X};\mu} + \boldsymbol{\zeta}_{\mu}'\boldsymbol{t}_{\mu}), \tag{A.1.16}$$

and covariance (1.18)

$$\boldsymbol{\sigma}_{\boldsymbol{X}}^{2} = -\frac{1}{2}(\underline{\boldsymbol{\theta}}_{\boldsymbol{X};\sigma,\sigma} + \boldsymbol{\zeta}_{\sigma}'\boldsymbol{t}_{\sigma,\sigma}\boldsymbol{\zeta}_{\sigma})^{-1}, \tag{A.1.17}$$

we can write the difference in (A.1.15) as follows

$$\begin{aligned}
\psi^N(\begin{pmatrix} \underline{\boldsymbol{\theta}}_{\boldsymbol{X};\mu} + \boldsymbol{\zeta}'_\mu \boldsymbol{t}_\mu \\ vec(\underline{\boldsymbol{\theta}}_{\boldsymbol{X};\sigma,\sigma} + \boldsymbol{\zeta}'_\sigma \boldsymbol{t}_{\sigma,\sigma}\boldsymbol{\zeta}_\sigma) \end{pmatrix}) - \psi^N(\underline{\boldsymbol{\theta}}_{\boldsymbol{X}}) &= \frac{1}{2}\boldsymbol{\mu}'_{\boldsymbol{X}}(\boldsymbol{\sigma}^2_{\boldsymbol{X}})^{-1}\boldsymbol{\mu}_{\boldsymbol{X}} - \frac{1}{2}\ln\det((\boldsymbol{\sigma}^2_{\boldsymbol{X}})^{-1}) \\
&\quad -\frac{1}{2}\underline{\boldsymbol{\mu}}'_{\boldsymbol{X}}(\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}})^{-1}\underline{\boldsymbol{\mu}}_{\boldsymbol{X}} + \frac{1}{2}\ln\det((\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}})^{-1}) \\
&= \frac{1}{2}[\boldsymbol{\mu}'_{\boldsymbol{X}}(\boldsymbol{\sigma}^2_{\boldsymbol{X}})^{-1}\boldsymbol{\mu}_{\boldsymbol{X}} - \underline{\boldsymbol{\mu}}'_{\boldsymbol{X}}(\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}})^{-1}\underline{\boldsymbol{\mu}}_{\boldsymbol{X}}] \\
&\quad -\frac{1}{2}[\ln\det((\boldsymbol{\sigma}^2_{\boldsymbol{X}})^{-1}) - \ln\det((\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}})^{-1})] \\
&= \frac{1}{2}[\boldsymbol{\mu}'_{\boldsymbol{X}}(\boldsymbol{\sigma}^2_{\boldsymbol{X}})^{-1}\boldsymbol{\mu}_{\boldsymbol{X}} - \underline{\boldsymbol{\mu}}'_{\boldsymbol{X}}(\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}})^{-1}\underline{\boldsymbol{\mu}}_{\boldsymbol{X}}] \\
&\quad -\frac{1}{2}[\ln\det((\boldsymbol{\sigma}^2_{\boldsymbol{X}})^{-1}\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}})] \\
&= \frac{1}{2}\boldsymbol{\mu}'_{\boldsymbol{X}}(\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}})^{-1}\boldsymbol{\mu}_{\boldsymbol{X}} - \boldsymbol{\mu}'_{\boldsymbol{X}}(\boldsymbol{\zeta}'_\sigma \boldsymbol{t}_{\sigma,\sigma}\boldsymbol{\zeta}_\sigma)\boldsymbol{\mu}_{\boldsymbol{X}} \\
&\quad -\frac{1}{2}\underline{\boldsymbol{\mu}}'_{\boldsymbol{X}}(\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}})^{-1}\underline{\boldsymbol{\mu}}_{\boldsymbol{X}} - \frac{1}{2}\ln\det(\mathbb{I}_{\bar{n}} - 2(\boldsymbol{\zeta}'_\sigma \boldsymbol{t}_{\sigma,\sigma}\boldsymbol{\zeta}_\sigma)\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}})],
\end{aligned} \quad \text{(A.1.18)}$$

where in the last row we used the relationship between canonical coordinates and covariance (1.18)-(1.11)

$$\begin{aligned}
(\boldsymbol{\sigma}^2_{\boldsymbol{X}})^{-1} &= -2(\underline{\boldsymbol{\theta}}_{\boldsymbol{X};\sigma,\sigma} + \boldsymbol{\zeta}'_\sigma \boldsymbol{t}_{\sigma,\sigma}\boldsymbol{\zeta}_\sigma) \\
&= -2(-\frac{1}{2}(\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}})^{-1} + \boldsymbol{\zeta}'_\sigma \boldsymbol{t}_{\sigma,\sigma}\boldsymbol{\zeta}_\sigma) \\
&= (\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}})^{-1} - 2\boldsymbol{\zeta}'_\sigma \boldsymbol{t}_{\sigma,\sigma}\boldsymbol{\zeta}_\sigma.
\end{aligned} \quad \text{(A.1.19)}$$

Also, from $(\underline{\boldsymbol{\theta}}_{\boldsymbol{X};\sigma,\sigma})^{-1} = -2\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}$ (1.11) and the binomial inverse theorem [Magnus and Neudecker, 1979] the yet-to-be-defined covariance also reads

$$\begin{aligned}
\boldsymbol{\sigma}^2_{\boldsymbol{X}} &= -\frac{1}{2}(\underline{\boldsymbol{\theta}}_{\boldsymbol{X};\sigma,\sigma} + \boldsymbol{\zeta}'_\sigma \boldsymbol{t}_{\sigma,\sigma}\boldsymbol{\zeta}_\sigma)^{-1} \\
&= -\frac{1}{2}((\underline{\boldsymbol{\theta}}_{\boldsymbol{X};\sigma,\sigma})^{-1} - (\underline{\boldsymbol{\theta}}_{\boldsymbol{X};\sigma,\sigma})^{-1}\boldsymbol{\zeta}'_\sigma[(\boldsymbol{\zeta}_\sigma(\underline{\boldsymbol{\theta}}_{\boldsymbol{X};\sigma,\sigma})^{-1}\boldsymbol{\zeta}'_\sigma + (\boldsymbol{t}_{\sigma,\sigma})^{-1})^{-1}\boldsymbol{\zeta}_\sigma(\underline{\boldsymbol{\theta}}_{\boldsymbol{X};\sigma,\sigma})^{-1}] \\
&= \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} + \boldsymbol{p}_\sigma\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}},
\end{aligned} \quad \text{(A.1.20)}$$

where we defined the $\bar{n} \times \bar{n}$ matrix

$$\boldsymbol{p}_\sigma \equiv \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}\boldsymbol{\zeta}'_\sigma(\frac{1}{2}(\boldsymbol{t}_{\sigma,\sigma})^{-1} - \boldsymbol{\zeta}_\sigma\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}\boldsymbol{\zeta}'_\sigma)^{-1}\boldsymbol{\zeta}_\sigma; \quad \text{(A.1.21)}$$

from which follows, using $\underline{\boldsymbol{\mu}}_{\boldsymbol{X}} = \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}\underline{\boldsymbol{\theta}}_{\boldsymbol{X};\mu}$ (A.1.42), the yet-to-be-defined expectation in turn

$$\begin{aligned}
\boldsymbol{\mu}_{\boldsymbol{X}} &= \boldsymbol{\sigma}^2_{\boldsymbol{X}}(\underline{\boldsymbol{\theta}}_{\boldsymbol{X};\mu} + \boldsymbol{\zeta}'_\mu \boldsymbol{t}_\mu) \\
&= (\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} + \boldsymbol{p}_\sigma\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}})(\underline{\boldsymbol{\theta}}_{\boldsymbol{X};\mu} + \boldsymbol{\zeta}'_\mu \boldsymbol{t}_\mu) \\
&= \underline{\boldsymbol{\mu}}_{\boldsymbol{X}} + \boldsymbol{p}_\sigma\underline{\boldsymbol{\mu}}_{\boldsymbol{X}} + (\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} + \boldsymbol{p}_\sigma\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}})\boldsymbol{\zeta}'_\mu \boldsymbol{t}_\mu.
\end{aligned} \quad \text{(A.1.22)}$$

A.1.4 Normal MRE update: views on expectations

In the case of only views on expectations as in (1.19) the updated canonical coordinates corresponding to the covariance must be the same as the base counterpart

$$\bar{\boldsymbol{\theta}}_{\boldsymbol{X};\sigma} = \underline{\boldsymbol{\theta}}_{\boldsymbol{X};\sigma,\sigma}, \tag{A.1.23}$$

or $\boldsymbol{\theta}_{\sigma}^{view} = \mathbf{0}_{\bar{k}\times\bar{k}}$.

Then the gradient of the dual Lagrangian (29) reads

$$\nabla_{\boldsymbol{t}_\mu}\mathcal{L}(\boldsymbol{t}_\mu;\boldsymbol{\eta}_\mu^{view}) = \boldsymbol{\zeta}_\mu\mathbb{E}^{f_{\boldsymbol{X}}}\left\{\boldsymbol{X}\right\} - \boldsymbol{\eta}_\mu^{view}. \tag{A.1.24}$$

In particular, since we know that the updated distribution must be normal (1.15), from (A.1.23) the gradient (A.1.24) becomes

$$\nabla_{\boldsymbol{t}_\mu}\mathcal{L}(\boldsymbol{t}_\mu;\boldsymbol{\eta}_\mu^{view}) = -\frac{1}{2}\boldsymbol{\zeta}_\mu(\underline{\boldsymbol{\theta}}_{\boldsymbol{X};\sigma,\sigma})^{-1}(\underline{\boldsymbol{\theta}}_{\boldsymbol{X};\mu} + \boldsymbol{\zeta}_\mu'\boldsymbol{t}_\mu) - \boldsymbol{\eta}_\mu^{view}. \tag{A.1.25}$$

Now if we set the above to zero, we obtain an exactly identified linear system of equations in the $\bar{k}$ yet-to-be-determined Lagrange multipliers $\boldsymbol{t}_\mu$

$$-\frac{1}{2}\boldsymbol{\zeta}_\mu(\underline{\boldsymbol{\theta}}_{\boldsymbol{X};\sigma,\sigma})^{-1}(\underline{\boldsymbol{\theta}}_{\boldsymbol{X};\mu} + \boldsymbol{\zeta}_\mu'\boldsymbol{t}_\mu) - \boldsymbol{\eta}_\mu^{view} = \mathbf{0}_{\bar{k}\times 1}. \tag{A.1.26}$$

The above can be easily re-written as follows

$$\boldsymbol{t}_\mu = -(\boldsymbol{\zeta}_\mu(\underline{\boldsymbol{\theta}}_{\boldsymbol{X};\sigma,\sigma})^{-1}\underline{\boldsymbol{\theta}}_{\boldsymbol{X};\mu} + 2\boldsymbol{\eta}_\mu^{view}). \tag{A.1.27}$$

Then, as long as $\boldsymbol{\zeta}_\mu$ is a full rank $\bar{k}$ matrix, the optimal Lagrange multipliers is unique and reads

$$\begin{aligned}\boldsymbol{\theta}_\mu^{view} &= -(\boldsymbol{\zeta}_\mu(\underline{\boldsymbol{\theta}}_{\boldsymbol{X};\sigma,\sigma})^{-1}\boldsymbol{\zeta}_\mu')^{-1}(\boldsymbol{\zeta}_\mu(\underline{\boldsymbol{\theta}}_{\boldsymbol{X};\sigma,\sigma})^{-1}\underline{\boldsymbol{\theta}}_{\boldsymbol{X};\mu} + 2\boldsymbol{\eta}_\mu^{view}) \\ &= (\boldsymbol{\zeta}_\mu\underline{\boldsymbol{\sigma}}_{\boldsymbol{X}}^2\boldsymbol{\zeta}_\mu')^{-1}(\boldsymbol{\eta}_\mu^{view} - \boldsymbol{\zeta}_\mu\underline{\boldsymbol{\mu}}_{\boldsymbol{X}}),\end{aligned} \tag{A.1.28}$$

where the last row follows from the relationship between base normal parameters and canonical coordinates (1.11).

Finally the updated canonical coordinates becomes

$$\bar{\boldsymbol{\theta}}_{\boldsymbol{X}} = \begin{pmatrix} \underline{\boldsymbol{\theta}}_{\boldsymbol{X};\mu} + \boldsymbol{\zeta}_\mu'(\boldsymbol{\zeta}_\mu\underline{\boldsymbol{\sigma}}_{\boldsymbol{X}}^2\boldsymbol{\zeta}_\mu')^{-1}(\boldsymbol{\eta}_\mu^{view} - \boldsymbol{\zeta}_\mu\underline{\boldsymbol{\mu}}_{\boldsymbol{X}}) \\ vec(\underline{\boldsymbol{\theta}}_{\boldsymbol{X};\sigma,\sigma}) \end{pmatrix}, \tag{A.1.29}$$

and hence, using (A.1.22) and (A.1.20), easily follows expectation $\bar{\boldsymbol{\mu}}_{\boldsymbol{X}}$ (1.21) and covariance $\bar{\boldsymbol{\sigma}}_{\boldsymbol{X}}^2$ (1.23) in turn.

Note how the uniqueness of the solution is connected to the invertibility of the component $\boldsymbol{\zeta}_\mu(\underline{\boldsymbol{\theta}}_{\boldsymbol{X};\sigma,\sigma})^{-1}\boldsymbol{\zeta}_\mu'$. This is not surprising, since the Hessian of the dual Lagrangian (30) reads

$$\nabla^2_{\boldsymbol{t}_\mu,\boldsymbol{t}_\mu}\mathcal{L}(\boldsymbol{t}_\mu;\boldsymbol{\eta}_\mu^{view}) = \boldsymbol{\zeta}_\mu\underline{\boldsymbol{\sigma}}_{\boldsymbol{X}}^2\boldsymbol{\zeta}_\mu', \tag{A.1.30}$$

and hence it is positive definite if and only if $\boldsymbol{\zeta}_\mu(\underline{\boldsymbol{\theta}}_{\boldsymbol{X};\sigma,\sigma})^{-1}\boldsymbol{\zeta}_\mu'$ is negative definite.

A.1.5 Normal MRE update: views on expectations as projection

It is easy to verify that the operator $\mathfrak{P}_{\boldsymbol{\zeta}_\mu}[\boldsymbol{x}]$ (1.25) is a projector.

Indeed $\mathfrak{P}_{\boldsymbol{\zeta}_\mu}[\boldsymbol{x}]$ is a linear transformation of the form $\mathfrak{P}_{\boldsymbol{\zeta}_\mu}[\boldsymbol{x}] = \boldsymbol{p}_\mu \boldsymbol{x}$, where $\boldsymbol{p}_\mu \equiv (\mathbb{I}_{\bar{n}} - \boldsymbol{\zeta}_\mu^{\dagger\prime}\boldsymbol{\zeta}_\mu)$ is an $\bar{n} \times \bar{n}$ idempotent matrix

$$\begin{aligned} \boldsymbol{p}_\mu^2 &= (\mathbb{I}_{\bar{n}} - \boldsymbol{\zeta}_\mu^{\dagger\prime}\boldsymbol{\zeta}_\mu)(\mathbb{I}_{\bar{n}} - \boldsymbol{\zeta}_\mu^{\dagger\prime}\boldsymbol{\zeta}_\mu) \\ &= \mathbb{I}_{\bar{n}} - \boldsymbol{\zeta}_\mu^{\dagger\prime}\boldsymbol{\zeta}_\mu = \boldsymbol{p}_\mu, \end{aligned} \tag{A.1.31}$$

as follows because the pseudo-inverse $\boldsymbol{\zeta}_\mu^\dagger$ (1.22) satisfies

$$\boldsymbol{\zeta}_\mu \boldsymbol{\zeta}_\mu^{\dagger\prime} = \mathbb{I}_{\bar{k}}. \tag{A.1.32}$$

Then (1.26) simply follows because we have

$$\begin{aligned} \mathfrak{P}_{\boldsymbol{\zeta}_\mu}[\underline{\boldsymbol{\mu}}_{\boldsymbol{X}} - \boldsymbol{\zeta}_\mu^{\dagger\prime}\boldsymbol{\eta}_\mu^{view}] &= (\mathbb{I}_{\bar{n}} - \boldsymbol{\zeta}_\mu^{\dagger\prime}\boldsymbol{\zeta}_\mu)(\underline{\boldsymbol{\mu}}_{\boldsymbol{X}} - \boldsymbol{\zeta}_\mu^{\dagger\prime}\boldsymbol{\eta}_\mu^{view}) \\ &= \underline{\boldsymbol{\mu}}_{\boldsymbol{X}} + \boldsymbol{\zeta}_\mu^{\dagger\prime}(\boldsymbol{\eta}_\mu^{view} - \boldsymbol{\zeta}_\mu\underline{\boldsymbol{\mu}}_{\boldsymbol{X}}) - \boldsymbol{\zeta}_\mu^{\dagger\prime}\boldsymbol{\eta}_\mu^{view} \\ &= \bar{\boldsymbol{\mu}}_{\boldsymbol{X}} - \boldsymbol{\zeta}_\mu^{\dagger\prime}\boldsymbol{\eta}_\mu^{view}, \end{aligned} \tag{A.1.33}$$

where in the last row we used the expression for the updated expectation (1.21).

Moreover the $\mathfrak{P}_{\boldsymbol{\zeta}_\mu}[\boldsymbol{x}]$ is orthogonal with respect to the inner product $\langle \boldsymbol{x}, \boldsymbol{y}\rangle_{\boldsymbol{\omega}^2} \equiv \boldsymbol{x}'\boldsymbol{\omega}^2\boldsymbol{y}$ induced by the inverse base or updated covariance $\boldsymbol{\omega}^2 \equiv (\underline{\boldsymbol{\sigma}}_{\boldsymbol{X}}^2)^{-1} = (\bar{\boldsymbol{\sigma}}_{\boldsymbol{X}}^2)^{-1}$ (1.23).

Indeed if we define the complementary projector

$$\mathfrak{P}_{\boldsymbol{\zeta}_\mu}^c[\boldsymbol{y}] \equiv \mathbb{I}_{\bar{n}} - \mathfrak{P}_{\boldsymbol{\zeta}_\mu}[\boldsymbol{x}], \tag{A.1.34}$$

then for any given $\bar{n} \times 1$ vector $\boldsymbol{x}, \boldsymbol{y}$, we have

$$\langle \mathfrak{P}_{\boldsymbol{\zeta}_\mu}[\boldsymbol{x}], \mathfrak{P}_{\boldsymbol{\zeta}_\mu}^c[\boldsymbol{y}]\rangle_{(\underline{\boldsymbol{\sigma}}_{\boldsymbol{X}}^2)^{-1}} = \boldsymbol{x}'(\mathbb{I}_{\bar{n}} - \boldsymbol{\zeta}_\mu^{\dagger\prime}\boldsymbol{\zeta}_\mu)'(\underline{\boldsymbol{\sigma}}_{\boldsymbol{X}}^2)^{-1}(\boldsymbol{\zeta}_\mu^{\dagger\prime}\boldsymbol{\zeta}_\mu)\boldsymbol{y} = 0, \tag{A.1.35}$$

as follows because the pseudo-inverse $\boldsymbol{\zeta}_\mu^\dagger$ (1.22) satisfies

$$(\underline{\boldsymbol{\sigma}}_{\boldsymbol{X}}^2)^{-1}(\boldsymbol{\zeta}_\mu^{\dagger\prime}\boldsymbol{\zeta}_\mu) = \boldsymbol{\zeta}_\mu'(\boldsymbol{\zeta}_\mu\underline{\boldsymbol{\sigma}}_{\boldsymbol{X}}^2\boldsymbol{\zeta}_\mu')^{-1}\boldsymbol{\zeta}_\mu \tag{A.1.36}$$

which implies

$$\begin{aligned} (\mathbb{I}_{\bar{n}} - \boldsymbol{\zeta}_\mu^{\dagger\prime}\boldsymbol{\zeta}_\mu)'(\underline{\boldsymbol{\sigma}}_{\boldsymbol{X}}^2)^{-1}(\boldsymbol{\zeta}_\mu^{\dagger\prime}\boldsymbol{\zeta}_\mu) &= \boldsymbol{\zeta}_\mu'(\boldsymbol{\zeta}_\mu\underline{\boldsymbol{\sigma}}_{\boldsymbol{X}}^2\boldsymbol{\zeta}_\mu')^{-1}\boldsymbol{\zeta}_\mu - \boldsymbol{\zeta}_\mu'(\boldsymbol{\zeta}_\mu\underline{\boldsymbol{\sigma}}_{\boldsymbol{X}}^2\boldsymbol{\zeta}_\mu')^{-1}\boldsymbol{\zeta}_\mu\underline{\boldsymbol{\sigma}}_{\boldsymbol{X}}^2\boldsymbol{\zeta}_\mu'(\boldsymbol{\zeta}_\mu\underline{\boldsymbol{\sigma}}_{\boldsymbol{X}}^2\boldsymbol{\zeta}_\mu')^{-1}\boldsymbol{\zeta}_\mu \\ &= \boldsymbol{\zeta}_\mu'(\boldsymbol{\zeta}_\mu\underline{\boldsymbol{\sigma}}_{\boldsymbol{X}}^2\boldsymbol{\zeta}_\mu')^{-1}\boldsymbol{\zeta}_\mu - \boldsymbol{\zeta}_\mu'(\boldsymbol{\zeta}_\mu\underline{\boldsymbol{\sigma}}_{\boldsymbol{X}}^2\boldsymbol{\zeta}_\mu')^{-1}\boldsymbol{\zeta}_\mu = \boldsymbol{0}. \end{aligned} \tag{A.1.37}$$

A.1.6 Normal MRE update: views on second non-central moments

In the case of only views on non-central second moments as in (1.33) the updated canonical coordinates corresponding to the expectation must be the same as the base counterpart

$$\bar{\boldsymbol{\theta}}_{\boldsymbol{X};\mu} = \underline{\boldsymbol{\theta}}_{\boldsymbol{X};\mu}, \tag{A.1.38}$$

or $\boldsymbol{\theta}_{\mu}^{view} = \mathbf{0}_{\bar{k}\times 1}$.

Then the gradient of the dual Lagrangian (29) reads

$$\nabla_{\boldsymbol{t}_{\sigma,\sigma}}\mathcal{L}(\boldsymbol{t}_{\sigma,\sigma};\boldsymbol{\eta}_{\sigma,\sigma}^{view}) = \boldsymbol{\zeta}_{\sigma}\mathbb{E}^{f_{\boldsymbol{X}}}\left\{\boldsymbol{X}\boldsymbol{X}'\right\}\boldsymbol{\zeta}_{\sigma}' - \boldsymbol{\eta}_{\sigma,\sigma}^{view}. \tag{A.1.39}$$

In particular, since we know that the updated distribution must be normal (1.15), from (A.1.38) the gradient (A.1.39) becomes

$$\nabla_{\boldsymbol{t}_{\sigma,\sigma}}\mathcal{L}(\boldsymbol{t}_{\sigma,\sigma};\boldsymbol{\eta}_{\sigma,\sigma}^{view}) = -\frac{1}{2}\boldsymbol{\zeta}_{\sigma}(\boldsymbol{t}_{\sigma,\sigma})^{-1}\boldsymbol{\zeta}_{\sigma}' + \boldsymbol{\zeta}_{\sigma}\boldsymbol{\mu}_{\boldsymbol{X}}(\boldsymbol{\zeta}_{\sigma}\boldsymbol{\mu}_{\boldsymbol{X}})' - \boldsymbol{\eta}_{\sigma,\sigma}^{view}, \tag{A.1.40}$$

where we defined

$$\boldsymbol{\theta}_{\boldsymbol{X};\sigma} \equiv \underline{\boldsymbol{\theta}}_{\boldsymbol{X};\sigma,\sigma} + \boldsymbol{\zeta}_{\sigma}'\boldsymbol{t}_{\sigma,\sigma}\boldsymbol{\zeta}_{\sigma}, \tag{A.1.41}$$

and where the view-implied expectation $\boldsymbol{\zeta}_{\sigma}\boldsymbol{\mu}_{\boldsymbol{X}}$ also reads

$$\boldsymbol{\zeta}_{\sigma}\boldsymbol{\mu}_{\boldsymbol{X}} = -\frac{1}{2}\boldsymbol{\zeta}_{\sigma}(\boldsymbol{\theta}_{\boldsymbol{X};\sigma})^{-1}\underline{\boldsymbol{\theta}}_{\boldsymbol{X};\mu}. \tag{A.1.42}$$

Now, from the binomial inverse theorem [Magnus and Neudecker, 1979] we can easily invert (A.1.41) as follows

$$(\boldsymbol{\theta}_{\boldsymbol{X};\sigma})^{-1} = (\underline{\boldsymbol{\theta}}_{\boldsymbol{X};\sigma,\sigma})^{-1} - (\underline{\boldsymbol{\theta}}_{\boldsymbol{X};\sigma,\sigma})^{-1}\boldsymbol{\zeta}_{\sigma}'(\boldsymbol{\zeta}_{\sigma}(\underline{\boldsymbol{\theta}}_{\boldsymbol{X};\sigma,\sigma})^{-1}\boldsymbol{\zeta}_{\sigma} + (\boldsymbol{t}_{\sigma,\sigma})^{-1})^{-1}\boldsymbol{\zeta}_{\sigma}(\underline{\boldsymbol{\theta}}_{\boldsymbol{X};\sigma,\sigma})^{-1}. \tag{A.1.43}$$

Then, if we set (A.1.40) to zero, we obtain a system of equations in $(\boldsymbol{\theta}_{\boldsymbol{X};\sigma})^{-1}$

$$-\frac{1}{2}\boldsymbol{\zeta}_{\sigma}(\boldsymbol{\theta}_{\boldsymbol{X};\sigma})^{-1}\boldsymbol{\zeta}_{\sigma}' + \boldsymbol{\zeta}_{\sigma}\boldsymbol{\mu}_{\boldsymbol{X}}(\boldsymbol{\zeta}_{\sigma}\boldsymbol{\mu}_{\boldsymbol{X}})' - \boldsymbol{\eta}_{\sigma,\sigma}^{view} = \mathbf{0}_{\bar{k}\times 1}. \tag{A.1.44}$$

In particular, from (A.1.43), if we define

$$\boldsymbol{\alpha} \equiv \boldsymbol{\zeta}_{\sigma}(\underline{\boldsymbol{\theta}}_{\boldsymbol{X};\sigma,\sigma})^{-1}\boldsymbol{\zeta}_{\sigma}', \tag{A.1.45}$$

we can re-write the term $\boldsymbol{\zeta}_{\sigma}(\boldsymbol{\theta}_{\boldsymbol{X};\sigma})^{-1}\boldsymbol{\zeta}_{\sigma}'$ as follows

$$\boldsymbol{\zeta}_{\sigma}(\boldsymbol{\theta}_{\boldsymbol{X};\sigma})^{-1}\boldsymbol{\zeta}_{\sigma}' = \boldsymbol{\alpha} - \boldsymbol{\alpha}(\boldsymbol{\alpha} + (\boldsymbol{t}_{\sigma,\sigma})^{-1})^{-1}\boldsymbol{\alpha}, \tag{A.1.46}$$

which allows to re-write (A.1.43) as

$$\boldsymbol{\alpha}(\boldsymbol{\alpha} + (\boldsymbol{t}_{\sigma,\sigma})^{-1})^{-1}\boldsymbol{\alpha} = \boldsymbol{\alpha} + 2(\boldsymbol{\eta}_{\sigma,\sigma}^{view} - \boldsymbol{\zeta}_{\sigma}\boldsymbol{\mu}_{\boldsymbol{X}}(\boldsymbol{\zeta}_{\sigma}\boldsymbol{\mu}_{\boldsymbol{X}})'). \tag{A.1.47}$$

Then, as long as $\boldsymbol{\zeta}_\sigma$ is a full rank $\bar{k}$ matrix, we can invert the $\bar{k}\times\bar{k}$ matrix $\boldsymbol{\alpha}$ and obtain

$$(\boldsymbol{\alpha}+(\boldsymbol{t}_{\sigma,\sigma})^{-1})^{-1}=\boldsymbol{\alpha}^{-1}+2\boldsymbol{\alpha}^{-1}(\boldsymbol{\eta}^{view}_{\sigma,\sigma}-\boldsymbol{\zeta}_\sigma\boldsymbol{\mu}_{\boldsymbol{X}}(\boldsymbol{\zeta}_\sigma\boldsymbol{\mu}_{\boldsymbol{X}})')\boldsymbol{\alpha}^{-1}, \tag{A.1.48}$$

which means again from the binomial inverse theorem

$$\boldsymbol{\alpha}+(\boldsymbol{t}_{\sigma,\sigma})^{-1}=\boldsymbol{\alpha}-(\boldsymbol{\alpha}+\frac{1}{2}(\boldsymbol{\eta}^{view}_{\sigma,\sigma}-\boldsymbol{\zeta}_\sigma\boldsymbol{\mu}_{\boldsymbol{X}}(\boldsymbol{\zeta}_\sigma\boldsymbol{\mu}_{\boldsymbol{X}})')^{-1})^{-1}. \tag{A.1.49}$$

Then from the above we have

$$\begin{aligned}(\boldsymbol{t}_{\sigma,\sigma})^{-1} &= -(\boldsymbol{\alpha}+\frac{1}{2}(\boldsymbol{\eta}^{view}_{\sigma,\sigma}-\boldsymbol{\zeta}_\sigma\boldsymbol{\mu}_{\boldsymbol{X}}(\boldsymbol{\zeta}_\sigma\boldsymbol{\mu}_{\boldsymbol{X}})')^{-1})^{-1}\\ &= 2((\boldsymbol{\zeta}_\sigma\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}\boldsymbol{\zeta}'_\sigma)^{-1}-(\boldsymbol{\eta}^{view}_{\sigma,\sigma}-\boldsymbol{\zeta}_\sigma\boldsymbol{\mu}_{\boldsymbol{X}}(\boldsymbol{\zeta}_\sigma\boldsymbol{\mu}_{\boldsymbol{X}})')^{-1})^{-1},\end{aligned} \tag{A.1.50}$$

and hence

$$\boldsymbol{t}_{\sigma,\sigma}=\frac{1}{2}((\boldsymbol{\zeta}_\sigma\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}\boldsymbol{\zeta}'_\sigma)^{-1}-(\boldsymbol{\eta}^{view}_{\sigma,\sigma}-\boldsymbol{\zeta}_\sigma\boldsymbol{\mu}_{\boldsymbol{X}}(\boldsymbol{\zeta}_\sigma\boldsymbol{\mu}_{\boldsymbol{X}})')^{-1}), \tag{A.1.51}$$

as follows from (A.1.45) and the relationship between base normal parameters and canonical coordinates (1.11)

$$\boldsymbol{\alpha}=-2(\boldsymbol{\zeta}_\sigma\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}\boldsymbol{\zeta}'_\sigma). \tag{A.1.52}$$

Finally, from the yet-to-be-determined Lagrange multipliers $\boldsymbol{t}_{\sigma,\sigma}$ we can also deduce the yet-to-be-determined covariance $\boldsymbol{\sigma}^2_{\boldsymbol{X}}=\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}+\boldsymbol{p}_\sigma\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}$ (A.1.20) from the $\bar{n}\times\bar{n}$ matrix $\boldsymbol{p}_\sigma$ (A.1.21) which simplifies to

$$\begin{aligned}\boldsymbol{p}_\sigma &= \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}\boldsymbol{\zeta}'_\sigma(\boldsymbol{\zeta}_\sigma\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}\boldsymbol{\zeta}'_\sigma)^{-1}(\boldsymbol{\eta}^{view}_{\sigma,\sigma}-\boldsymbol{\zeta}_\sigma\boldsymbol{\mu}_{\boldsymbol{X}}(\boldsymbol{\zeta}_\sigma\boldsymbol{\mu}_{\boldsymbol{X}})'-\boldsymbol{\zeta}_\sigma\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}\boldsymbol{\zeta}'_\sigma)(\boldsymbol{\zeta}_\sigma\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}\boldsymbol{\zeta}'_\sigma)^{-1}\boldsymbol{\zeta}_\sigma\\ &= \boldsymbol{\zeta}^{\dagger\prime}_\sigma(\boldsymbol{\eta}^{view}_{\sigma,\sigma}-\boldsymbol{\zeta}_\sigma\boldsymbol{\mu}_{\boldsymbol{X}}(\boldsymbol{\zeta}_\sigma\boldsymbol{\mu}_{\boldsymbol{X}})'-\boldsymbol{\zeta}_\sigma\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}\boldsymbol{\zeta}'_\sigma)(\boldsymbol{\zeta}_\sigma\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}\boldsymbol{\zeta}'_\sigma)^{-1}\boldsymbol{\zeta}_\sigma,\end{aligned} \tag{A.1.53}$$

as follows from (A.1.51) and the binomial inverse theorem [Magnus and Neudecker, 1979], or

$$\begin{aligned}\frac{1}{2}(\boldsymbol{t}_{\sigma,\sigma})^{-1} &= ((\boldsymbol{\zeta}_\sigma\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}\boldsymbol{\zeta}'_\sigma)^{-1}-(\boldsymbol{\eta}^{view}_{\sigma,\sigma}-\boldsymbol{\zeta}_\sigma\boldsymbol{\mu}_{\boldsymbol{X}}(\boldsymbol{\zeta}_\sigma\boldsymbol{\mu}_{\boldsymbol{X}})')^{-1})^{-1}\\ &= \boldsymbol{\zeta}_\sigma\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}\boldsymbol{\zeta}'_\sigma+(\boldsymbol{\zeta}_\sigma\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}\boldsymbol{\zeta}'_\sigma)(\boldsymbol{\eta}^{view}_{\sigma,\sigma}-\boldsymbol{\zeta}_\sigma\boldsymbol{\mu}_{\boldsymbol{X}}(\boldsymbol{\zeta}_\sigma\boldsymbol{\mu}_{\boldsymbol{X}})'-\boldsymbol{\zeta}_\sigma\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}\boldsymbol{\zeta}'_\sigma)^{-1}(\boldsymbol{\zeta}_\sigma\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}\boldsymbol{\zeta}'_\sigma).\end{aligned} \tag{A.1.54}$$

Then, the yet-to-be-determined expectation $\boldsymbol{\mu}_{\boldsymbol{X}}=\underline{\boldsymbol{\mu}}_{\boldsymbol{X}}+\boldsymbol{p}_\sigma\underline{\boldsymbol{\mu}}_{\boldsymbol{X}}$ (A.1.22) follows accordingly.

Moreover, it is easy to verify that the following identity holds

$$\begin{aligned}\boldsymbol{\zeta}_\sigma\boldsymbol{\mu}_{\boldsymbol{X}} &= \boldsymbol{\zeta}_\sigma\underline{\boldsymbol{\mu}}_{\boldsymbol{X}}+\boldsymbol{\zeta}_\sigma\boldsymbol{p}_\sigma\underline{\boldsymbol{\mu}}_{\boldsymbol{X}}\\ &= \boldsymbol{\zeta}_\sigma\underline{\boldsymbol{\mu}}_{\boldsymbol{X}}+(\boldsymbol{\eta}^{view}_{\sigma,\sigma}-\boldsymbol{\zeta}_\sigma\boldsymbol{\mu}_{\boldsymbol{X}}(\boldsymbol{\zeta}_\sigma\boldsymbol{\mu}_{\boldsymbol{X}})'-\boldsymbol{\zeta}_\sigma\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}\boldsymbol{\zeta}'_\sigma)(\boldsymbol{\zeta}_\sigma\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}\boldsymbol{\zeta}'_\sigma)^{-1}\boldsymbol{\zeta}_\sigma\underline{\boldsymbol{\mu}}_{\boldsymbol{X}}\\ &= (\boldsymbol{\eta}^{view}_{\sigma,\sigma}-\boldsymbol{\zeta}_\sigma\boldsymbol{\mu}_{\boldsymbol{X}}(\boldsymbol{\zeta}_\sigma\boldsymbol{\mu}_{\boldsymbol{X}})')(\boldsymbol{\zeta}_\sigma\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}\boldsymbol{\zeta}'_\sigma)^{-1}\boldsymbol{\zeta}_\sigma\underline{\boldsymbol{\mu}}_{\boldsymbol{X}},\end{aligned} \tag{A.1.55}$$

as follows from

$$\begin{aligned}\boldsymbol{\mu}_{\boldsymbol{X}} &= \underline{\boldsymbol{\mu}}_{\boldsymbol{X}}+\boldsymbol{p}_\sigma\underline{\boldsymbol{\mu}}_{\boldsymbol{X}}\\ &= \underline{\boldsymbol{\mu}}_{\boldsymbol{X}}+\boldsymbol{\zeta}^{\dagger\prime}_\sigma((\boldsymbol{\eta}^{view}_{\sigma,\sigma}-\boldsymbol{\zeta}_\sigma\boldsymbol{\mu}_{\boldsymbol{X}}(\boldsymbol{\zeta}_\sigma\boldsymbol{\mu}_{\boldsymbol{X}})')(\boldsymbol{\zeta}_\sigma\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}\boldsymbol{\zeta}'_\sigma)^{-1}\boldsymbol{\zeta}_\sigma\underline{\boldsymbol{\mu}}_{\boldsymbol{X}}-\boldsymbol{\zeta}_\sigma\underline{\boldsymbol{\mu}}_{\boldsymbol{X}}).\end{aligned} \tag{A.1.56}$$

To conclude, to find the optimal Lagrange multipliers $\boldsymbol{t}_{\sigma,\sigma}$ we need to solve an implicit system of equations. This can be done numerically via recursion, as explained in Table 1.2.

A.1.7 Normal MRE update: views on covariance as projection

First of all, it is easy to verify that the pseudo-inverse $\zeta_\sigma^\dagger \equiv (\zeta_\sigma \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} \zeta'_\sigma)^{-1} \zeta_\sigma \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}$ satisfies (1.39).

Indeed, from the expression of the updated covariance $\bar{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}$ (1.38), we have

$$
\begin{aligned}
(\zeta_\sigma \bar{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} \zeta'_\sigma)^{-1} \zeta_\sigma \bar{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} &= (\boldsymbol{\sigma}^{2view})^{-1} \zeta_\sigma (\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} + \zeta_\sigma^{\dagger\prime} (\boldsymbol{\sigma}^{2view} - \zeta_\sigma \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} \zeta'_\sigma) \zeta_\sigma^\dagger) \\
&= (\boldsymbol{\sigma}^{2view})^{-1} \zeta_\sigma \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} + \zeta_\sigma^\dagger - (\boldsymbol{\sigma}^{2view})^{-1} \zeta_\sigma \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} \\
&= \zeta_\sigma^\dagger,
\end{aligned} \tag{A.1.57}
$$

where the second row follows because $\zeta_\sigma \zeta_\sigma^{\dagger\prime} = \mathbb{I}_{\bar{k}}$.

Then, following similar arguments as in [A.1.5], we obtain that $\mathfrak{P}_{\zeta_\sigma}[\boldsymbol{x}]$ (1.56) is an orthogonal projector with respect to the inner product $\langle \boldsymbol{x}, \boldsymbol{y} \rangle_{\boldsymbol{\omega}^2} \equiv \boldsymbol{x}' \boldsymbol{\omega}^2 \boldsymbol{y}$ induced by either i) the inverse base covariance $\boldsymbol{\omega}^2 \equiv (\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}})^{-1}$, or ii) the inverse updated covariance $\boldsymbol{\omega}^2 \equiv (\bar{\boldsymbol{\sigma}}^2_{\boldsymbol{X}})^{-1}$ (1.38).

Also, it turns out that the pseudo-inverse $\zeta_\sigma^\dagger \equiv (\zeta_\sigma \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} \zeta'_\sigma)^{-1} \zeta_\sigma \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}$ satisfies

$$
\begin{aligned}
\zeta_\sigma^\dagger \otimes \zeta_\sigma^\dagger &= ((\zeta_\sigma \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} \zeta'_\sigma)^{-1} \otimes (\zeta_\sigma \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} \zeta'_\sigma)^{-1})(\zeta_\sigma \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} \otimes \zeta_\sigma \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}) \\
&= (\zeta_\sigma \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} \zeta'_\sigma \otimes \zeta_\sigma \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} \zeta'_\sigma)^{-1} (\zeta_\sigma \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} \otimes \zeta_\sigma \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}) \\
&= ((\zeta_\sigma \otimes \zeta_\sigma)(\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} \zeta'_\sigma \otimes \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} \zeta'_\sigma))^{-1} (\zeta_\sigma \otimes \zeta_\sigma)(\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} \otimes \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}) \\
&= ((\zeta_\sigma \otimes \zeta_\sigma)(\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} \otimes \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}})(\zeta_\sigma \otimes \zeta_\sigma)')^{-1} (\zeta_\sigma \otimes \zeta_\sigma)(\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} \otimes \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}) \\
&\equiv (\zeta_\sigma \otimes \zeta_\sigma)^\dagger,
\end{aligned} \tag{A.1.58}
$$

as follows from the properties of the Kronecker product, see [Magnus and Neudecker, 1979]. Note that according to (1.39), the pseudo inverse $(\zeta_\sigma \otimes \zeta_\sigma)^\dagger$ can be also equivalently weighted via the Kronecker product of updated covariances $(\bar{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} \otimes \bar{\boldsymbol{\sigma}}^2_{\boldsymbol{X}})$.

This implies that the vectorized updated covariance $\bar{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}$ (1.38) satisfies (1.58).

Indeed we have

$$
\begin{aligned}
vec(\bar{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}) &= vec(\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}) + vec(\zeta_\sigma^{\dagger\prime} (\boldsymbol{\sigma}^{2view} - \zeta_\sigma \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} \zeta'_\sigma) \zeta_\sigma^\dagger) \\
&= vec(\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}) + (\zeta_\sigma^\dagger \otimes \zeta_\sigma^\dagger)' vec(\boldsymbol{\sigma}^{2view} - \zeta_\sigma \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} \zeta'_\sigma) \\
&= vec(\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}) + (\zeta_\sigma \otimes \zeta_\sigma)^{\dagger\prime} vec(\boldsymbol{\sigma}^{2view} - \zeta_\sigma \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} \zeta'_\sigma),
\end{aligned} \tag{A.1.59}
$$

as follows from the properties of the Kronecker product, see [Magnus and Neudecker, 1979], and (A.1.58).

Then since (A.1.59) is of the same form as the updated expectation (1.37), we can use similar arguments to obtain (1.58).

A.1.8 Normal MRE update: views on expectations and second non-central moments

In the case of joint views on expectations and second non-central moments as in (1.9) the gradient of the dual Lagrangian (29) reads

$$\begin{aligned}\nabla_{\boldsymbol{t}}\mathcal{L}(\boldsymbol{t};\boldsymbol{\eta}^{view}) &\equiv \begin{pmatrix}\nabla_{\boldsymbol{t}_\mu}\mathcal{L}(\boldsymbol{t}_\mu;\boldsymbol{\eta}_\mu^{view})\\ vec(\nabla_{\boldsymbol{t}_{\sigma,\sigma}}\mathcal{L}(\boldsymbol{t}_{\sigma,\sigma};\boldsymbol{\eta}_{\sigma,\sigma}^{view}))\end{pmatrix} \\ &= \begin{pmatrix}\boldsymbol{\zeta}_\mu\mathbb{E}^{f_{\boldsymbol{X}}}\{\boldsymbol{X}\}-\boldsymbol{\eta}_\mu^{view}\\ vec(\boldsymbol{\zeta}_\sigma\mathbb{E}^{f_{\boldsymbol{X}}}\{\boldsymbol{X}\boldsymbol{X}'\}\boldsymbol{\zeta}'_\sigma-\boldsymbol{\eta}_{\sigma,\sigma}^{view})\end{pmatrix},\end{aligned} \tag{A.1.60}$$

Then, following similar arguments as in Appendix A.1.4 and A.1.6, it is easy to verify the following holds true

$$\nabla_{\boldsymbol{t}_\mu}\mathcal{L}(\boldsymbol{t}_\mu;\boldsymbol{\eta}_\mu^{view}) = \boldsymbol{\zeta}_\mu\boldsymbol{\sigma}^2_{\boldsymbol{X}}(\underline{\boldsymbol{\theta}}_{\boldsymbol{X};\mu}+\boldsymbol{\zeta}'_\mu\boldsymbol{t}_\mu)-\boldsymbol{\eta}_\mu^{view}, \tag{A.1.61}$$

and

$$\nabla_{\boldsymbol{t}_{\sigma,\sigma}}\mathcal{L}(\boldsymbol{t}_{\sigma,\sigma};\boldsymbol{\eta}_{\sigma,\sigma}^{view}) = -\frac{1}{2}\boldsymbol{\zeta}_\sigma(\underline{\boldsymbol{\theta}}_{\boldsymbol{X};\sigma,\sigma}+\boldsymbol{\zeta}'_\sigma\boldsymbol{t}_{\sigma,\sigma}\boldsymbol{\zeta}_\sigma)^{-1}\boldsymbol{\zeta}'_\sigma+\boldsymbol{\zeta}_\sigma\boldsymbol{\mu}_{\boldsymbol{X}}(\boldsymbol{\zeta}_\sigma\boldsymbol{\mu}_{\boldsymbol{X}})'-\boldsymbol{\eta}_{\sigma,\sigma}^{view}, \tag{A.1.62}$$

where $\boldsymbol{\sigma}^2_{\boldsymbol{X}} = \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} + \boldsymbol{p}_\sigma\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}$ is the yet-to-be-determined covariance as in (A.1.20) and $\boldsymbol{\zeta}_\sigma\boldsymbol{\mu}_{\boldsymbol{X}}$ follows from (A.1.22)

$$\boldsymbol{\eta}_\sigma^{view} \equiv \boldsymbol{\zeta}_\sigma\boldsymbol{\mu}_{\boldsymbol{X}} \equiv -\frac{1}{2}\boldsymbol{\zeta}_\sigma(\underline{\boldsymbol{\theta}}_{\boldsymbol{X};\sigma,\sigma}+\boldsymbol{\zeta}'_\sigma\boldsymbol{t}_{\sigma,\sigma}\boldsymbol{\zeta}_\sigma)^{-1}(\underline{\boldsymbol{\theta}}_{\boldsymbol{X};\mu}+\boldsymbol{\zeta}'_\mu\boldsymbol{t}_\mu). \tag{A.1.63}$$

In particular, if we set (A.1.61) to zero and follow similar arguments as in Appendix A.1.4, the optimal Lagrange multipliers corresponding to the views on expectations must satisfy the following

$$\boldsymbol{t}_\mu = (\boldsymbol{\zeta}_\mu\boldsymbol{\sigma}^2_{\boldsymbol{X}}\boldsymbol{\zeta}'_\mu)^{-1}(\boldsymbol{\eta}_\mu^{view}-\boldsymbol{\zeta}_\mu\boldsymbol{\sigma}^2_{\boldsymbol{X}}(\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}})^{-1}\underline{\boldsymbol{\mu}}_{\boldsymbol{X}}). \tag{A.1.64}$$

From the other hand, if we set (A.1.62) to zero and follow similar computations as in Appendix A.1.6, the optimal Lagrange multipliers corresponding to the views on second non-central moments must satisfy the following

$$\boldsymbol{t}_{\sigma,\sigma} = \frac{1}{2}((\boldsymbol{\zeta}_\sigma\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}\boldsymbol{\zeta}'_\sigma)^{-1}-(\boldsymbol{\eta}_{\sigma,\sigma}^{view}-\boldsymbol{\zeta}_\sigma\boldsymbol{\mu}_{\boldsymbol{X}}(\boldsymbol{\zeta}_\sigma\boldsymbol{\mu}_{\boldsymbol{X}})')^{-1}). \tag{A.1.65}$$

This implies that the yet-to-be-determined expectation (A.1.22) becomes

$$\begin{aligned}\boldsymbol{\mu}_{\boldsymbol{X}} &= \boldsymbol{\sigma}^2_{\boldsymbol{X}}(\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}})^{-1}\underline{\boldsymbol{\mu}}_{\boldsymbol{X}}+\boldsymbol{\sigma}^2_{\boldsymbol{X}}\boldsymbol{\zeta}'_\mu(\boldsymbol{\zeta}_\mu\boldsymbol{\sigma}^2_{\boldsymbol{X}}\boldsymbol{\zeta}'_\mu)^{-1}(\boldsymbol{\eta}_\mu^{view}-\boldsymbol{\zeta}_\mu\boldsymbol{\sigma}^2_{\boldsymbol{X}}(\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}})^{-1}\underline{\boldsymbol{\mu}}_{\boldsymbol{X}}) \\ &= \boldsymbol{\mu}_{\boldsymbol{X};\sigma}+\boldsymbol{\sigma}^2_{\boldsymbol{X}}\boldsymbol{\zeta}'_\mu(\boldsymbol{\zeta}_\mu\boldsymbol{\sigma}^2_{\boldsymbol{X}}\boldsymbol{\zeta}'_\mu)^{-1}(\boldsymbol{\eta}_\mu^{view}-\boldsymbol{\zeta}_\mu\boldsymbol{\mu}_{\boldsymbol{X};\sigma}),\end{aligned} \tag{A.1.66}$$

as follows from similar arguments as in [A.1.6] and

$$\begin{aligned}\boldsymbol{\sigma}^2_{\boldsymbol{X}}(\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}})^{-1}\underline{\boldsymbol{\mu}}_{\boldsymbol{X}} &= (\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}+\boldsymbol{p}_\sigma\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}})(\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}})^{-1}\underline{\boldsymbol{\mu}}_{\boldsymbol{X}} \\ &= \underline{\boldsymbol{\mu}}_{\boldsymbol{X}}+\boldsymbol{p}_\sigma\underline{\boldsymbol{\mu}}_{\boldsymbol{X}}.\end{aligned} \tag{A.1.67}$$

Finally, the following identity holds

$$\boldsymbol{\zeta}_\sigma \boldsymbol{\mu}_{\boldsymbol{X}} = \boldsymbol{\zeta}_\sigma(\underline{\boldsymbol{\mu}}_{\boldsymbol{X}} + \boldsymbol{p}_\sigma \underline{\boldsymbol{\mu}}_{\boldsymbol{X}}) + \boldsymbol{\zeta}_\sigma \boldsymbol{\sigma}^2_{\boldsymbol{X}} \boldsymbol{\zeta}'_\mu (\boldsymbol{\zeta}_\mu \boldsymbol{\sigma}^2_{\boldsymbol{X}} \boldsymbol{\zeta}'_\mu)^{-1}(\boldsymbol{\eta}^{view}_\mu - \boldsymbol{\zeta}_\mu(\underline{\boldsymbol{\mu}}_{\boldsymbol{X}} + \boldsymbol{p}_\sigma \underline{\boldsymbol{\mu}}_{\boldsymbol{X}})), \tag{A.1.68}$$

where, following similar arguments as in [A.1.6], we have

$$\begin{aligned} \underline{\boldsymbol{\mu}}_{\boldsymbol{X}} + \boldsymbol{p}_\sigma \underline{\boldsymbol{\mu}}_{\boldsymbol{X}} &= \underline{\boldsymbol{\mu}}_{\boldsymbol{X}} + \boldsymbol{\zeta}^{\dagger\prime}_\sigma((\boldsymbol{\eta}^{view}_{\sigma,\sigma} - \boldsymbol{\zeta}_\sigma \boldsymbol{\mu}_{\boldsymbol{X}}(\boldsymbol{\zeta}_\sigma \boldsymbol{\mu}_{\boldsymbol{X}})')(\boldsymbol{\zeta}_\sigma \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} \boldsymbol{\zeta}'_\sigma)^{-1} \boldsymbol{\zeta}_\sigma \underline{\boldsymbol{\mu}}_{\boldsymbol{X}} - \boldsymbol{\zeta}_\sigma \underline{\boldsymbol{\mu}}_{\boldsymbol{X}}) \\ &= \boldsymbol{\mu}_{\boldsymbol{X};\sigma}, \end{aligned} \tag{A.1.69}$$

which implies also

$$\begin{aligned} \boldsymbol{\zeta}_\sigma \boldsymbol{\mu}_{\boldsymbol{X};\sigma} &= \boldsymbol{\zeta}_\sigma(\underline{\boldsymbol{\mu}}_{\boldsymbol{X}} + \boldsymbol{p}_\sigma \underline{\boldsymbol{\mu}}_{\boldsymbol{X}}) \\ &= (\boldsymbol{\eta}^{view}_{\sigma,\sigma} - \boldsymbol{\zeta}_\sigma \boldsymbol{\mu}_{\boldsymbol{X}}(\boldsymbol{\zeta}_\sigma \boldsymbol{\mu}_{\boldsymbol{X}})')(\boldsymbol{\zeta}_\sigma \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} \boldsymbol{\zeta}'_\sigma)^{-1} \boldsymbol{\zeta}_\sigma \underline{\boldsymbol{\mu}}_{\boldsymbol{X}}. \end{aligned} \tag{A.1.70}$$

To conclude, to find the optimal Lagrange multipliers $(\boldsymbol{t}_\mu, \boldsymbol{t}_{\sigma,\sigma})$ we need to solve an implicit system of equations. This can be done numerically via recursion, as explained in Table 1.5.

A.1.9 Normal MRE update: same view variables

Consider joint equality views (18) on *same* linear combinations as in (1.72).

In this case, the updated expectation of the view variables $\boldsymbol{\eta}^{view}_\sigma$ (1.34) becomes explicit (1.73). This means that formulas in Table 1.6 can be easily recovered by replacing

$$\boldsymbol{\eta}^{view}_\sigma = \boldsymbol{\eta}^{view}_\mu, \tag{A.1.71}$$

in the respective general MRE solutions (1.61)-(1.63)-(1.64)-(1.66).

Namely, it is immediate that the updated covariance (1.66) becomes

$$\bar{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} = \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} + \boldsymbol{\zeta}^{\dagger\prime}(\boldsymbol{\eta}^{view}_{\sigma,\sigma} - \boldsymbol{\eta}^{view}_\mu \boldsymbol{\eta}^{view\prime}_\mu - \boldsymbol{\zeta}\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} \boldsymbol{\zeta}')\boldsymbol{\zeta}^\dagger, \tag{A.1.72}$$

since by construction we have

$$\sigma^{2view}(\boldsymbol{\eta}^{view}_\sigma) = \boldsymbol{\eta}^{view}_{\sigma,\sigma} - \boldsymbol{\eta}^{view}_\mu \boldsymbol{\eta}^{view\prime}_\mu. \tag{A.1.73}$$

Then, the optimal Lagrange multipliers $\boldsymbol{\theta}^{view}_{\sigma,\sigma}$ (1.63) reads

$$\boldsymbol{\theta}^{view}_{\sigma,\sigma} = \frac{1}{2}((\boldsymbol{\zeta}\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} \boldsymbol{\zeta}')^{-1} - (\boldsymbol{\eta}^{view}_{\sigma,\sigma} - \boldsymbol{\eta}^{view}_\mu \boldsymbol{\eta}^{view\prime}_\mu)^{-1}); \tag{A.1.74}$$

and the optimal Lagrange multipliers $\boldsymbol{\theta}^{view}_\mu$ (1.61) reads

$$\boldsymbol{\theta}^{view}_\mu = (\boldsymbol{\eta}^{view}_{\sigma,\sigma} - \boldsymbol{\eta}^{view}_\mu \boldsymbol{\eta}^{view\prime}_\mu)^{-1} \boldsymbol{\eta}^{view}_\mu - (\boldsymbol{\zeta}\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} \boldsymbol{\zeta}')^{-1} \boldsymbol{\zeta}\underline{\boldsymbol{\mu}}_{\boldsymbol{X}}, \tag{A.1.75}$$

as follows because

$$\bar{\boldsymbol{\mu}}_{\boldsymbol{X};\sigma} = \underline{\boldsymbol{\mu}}_{\boldsymbol{X}} + \boldsymbol{\zeta}^{\dagger\prime}((\boldsymbol{\eta}^{view}_{\sigma,\sigma} - \boldsymbol{\eta}^{view}_\mu \boldsymbol{\eta}^{view\prime}_\mu)(\boldsymbol{\zeta}\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} \boldsymbol{\zeta}')^{-1} \boldsymbol{\zeta}\underline{\boldsymbol{\mu}}_{\boldsymbol{X}} - \boldsymbol{\zeta}\underline{\boldsymbol{\mu}}_{\boldsymbol{X}}), \tag{A.1.76}$$

and hence

$$\boldsymbol{\zeta}\bar{\boldsymbol{\mu}}_{\boldsymbol{X};\sigma} = (\boldsymbol{\eta}_{\sigma,\sigma}^{view} - \boldsymbol{\eta}_{\mu}^{view}\boldsymbol{\eta}_{\mu}^{view\prime})(\boldsymbol{\zeta}\underline{\boldsymbol{\sigma}}_{\boldsymbol{X}}^2\boldsymbol{\zeta}')^{-1}\boldsymbol{\zeta}\underline{\boldsymbol{\mu}}_{\boldsymbol{X}}. \tag{A.1.77}$$

Finally the updated expectation (1.64) easily follows from (1.64)

$$\begin{aligned}
\bar{\boldsymbol{\mu}}_{\boldsymbol{X}} &= \bar{\boldsymbol{\mu}}_{\boldsymbol{X};\sigma} + \boldsymbol{\zeta}^{\dagger\prime}(\boldsymbol{\eta}_{\mu}^{view} - \boldsymbol{\zeta}\bar{\boldsymbol{\mu}}_{\boldsymbol{X};\sigma}) \\
&= \underline{\boldsymbol{\mu}}_{\boldsymbol{X}} + \boldsymbol{\zeta}^{\dagger\prime}((\boldsymbol{\eta}_{\sigma,\sigma}^{view} - \boldsymbol{\eta}_{\mu}^{view}\boldsymbol{\eta}_{\mu}^{view\prime})(\boldsymbol{\zeta}\underline{\boldsymbol{\sigma}}_{\boldsymbol{X}}^2\boldsymbol{\zeta}')^{-1}\boldsymbol{\zeta}\underline{\boldsymbol{\mu}}_{\boldsymbol{X}} - \boldsymbol{\zeta}\underline{\boldsymbol{\mu}}_{\boldsymbol{X}}) \\
&\quad + \boldsymbol{\zeta}^{\dagger\prime}(\boldsymbol{\eta}_{\mu}^{view} - (\boldsymbol{\eta}_{\sigma,\sigma}^{view} - \boldsymbol{\eta}_{\mu}^{view}\boldsymbol{\eta}_{\mu}^{view\prime})(\boldsymbol{\zeta}\underline{\boldsymbol{\sigma}}_{\boldsymbol{X}}^2\boldsymbol{\zeta}')^{-1}\boldsymbol{\zeta}\underline{\boldsymbol{\mu}}_{\boldsymbol{X}}) \\
&= \underline{\boldsymbol{\mu}}_{\boldsymbol{X}} + \boldsymbol{\zeta}^{\dagger\prime}(\boldsymbol{\eta}_{\mu}^{view} - \boldsymbol{\zeta}\underline{\boldsymbol{\mu}}_{\boldsymbol{X}}).
\end{aligned} \tag{A.1.78}$$

A.1.10 Normal MRE update: views on expectations and covariances

Consider joint equality views (18) on expectation and covariance as in (1.80).

In this case, the updated covariance of the view variables $\sigma^{2view}(\boldsymbol{\eta}_{\sigma}^{view})$ (1.35) becomes explicit (1.82). This means that formulas in Table 1.7 can be easily recovered by replacing

$$\sigma^{2view}(\boldsymbol{\eta}_{\sigma}^{view}) = \boldsymbol{\sigma}^{2view}, \tag{A.1.79}$$

in the respective general MRE solutions (1.61)-(1.63)-(1.64)-(1.66).

Namely, it is immediate that the updated covariance (1.66) becomes

$$\bar{\boldsymbol{\sigma}}_{\boldsymbol{X}}^2 = \underline{\boldsymbol{\sigma}}_{\boldsymbol{X}}^2 + \zeta_{\sigma}^{\dagger\prime}(\boldsymbol{\sigma}^{2view} - \zeta_{\sigma}\underline{\boldsymbol{\sigma}}_{\boldsymbol{X}}^2\zeta_{\sigma}')\zeta_{\sigma}^{\dagger}, \tag{A.1.80}$$

Then, the optimal Lagrange multipliers $\boldsymbol{\theta}_{\sigma,\sigma}^{view}$ (1.63) reads

$$\boldsymbol{\theta}_{\sigma,\sigma}^{view} = \frac{1}{2}((\boldsymbol{\zeta}\underline{\boldsymbol{\sigma}}_{\boldsymbol{X}}^2\boldsymbol{\zeta}')^{-1} - (\boldsymbol{\sigma}^{2view})^{-1}); \tag{A.1.81}$$

and the optimal Lagrange multipliers $\boldsymbol{\theta}_{\mu}^{view}$ (1.61) reads

$$\boldsymbol{\theta}_{\mu}^{view} = (\boldsymbol{\zeta}_{\mu}\bar{\boldsymbol{\sigma}}_{\boldsymbol{X}}^2\boldsymbol{\zeta}_{\mu}')^{-1}(\boldsymbol{\mu}^{view} - \boldsymbol{\zeta}_{\mu}\bar{\boldsymbol{\mu}}_{\boldsymbol{X};\sigma}), \tag{A.1.82}$$

where

$$\bar{\boldsymbol{\mu}}_{\boldsymbol{X};\sigma} = \underline{\boldsymbol{\mu}}_{\boldsymbol{X}} + \zeta_{\sigma}^{\dagger\prime}(\boldsymbol{\sigma}^{2view}(\zeta_{\sigma}\underline{\boldsymbol{\sigma}}_{\boldsymbol{X}}^2\zeta_{\sigma}')^{-1}\zeta_{\sigma}\underline{\boldsymbol{\mu}}_{\boldsymbol{X}} - \zeta_{\sigma}\underline{\boldsymbol{\mu}}_{\boldsymbol{X}}), \tag{A.1.83}$$

and similar follows for the updated expectation $\bar{\boldsymbol{\mu}}_{\boldsymbol{X}}$.

Finally, it turns out that if view variables are statistically independent $\boldsymbol{\zeta}_{\mu}\underline{\boldsymbol{\sigma}}_{\boldsymbol{X}}^2\zeta_{\sigma}'$ (1.89), then

$$\boldsymbol{\zeta}_{\mu}\bar{\boldsymbol{\mu}}_{\boldsymbol{X};\sigma} = \boldsymbol{\zeta}_{\mu}\underline{\boldsymbol{\mu}}_{\boldsymbol{X}}, \tag{A.1.84}$$

as follows because $\boldsymbol{\zeta}_{\mu}\zeta_{\sigma}^{\dagger\prime} = \boldsymbol{\zeta}_{\mu}\underline{\boldsymbol{\sigma}}_{\boldsymbol{X}}^2\zeta_{\sigma}'(\zeta_{\sigma}\underline{\boldsymbol{\sigma}}_{\boldsymbol{X}}^2\zeta_{\sigma}')^{-1} = \mathbf{0}$, which implies the optimal Lagrange multipliers $\boldsymbol{\theta}_{\mu}^{view}$ (1.61)

$$\boldsymbol{\theta}_{\mu}^{view} = (\boldsymbol{\zeta}_{\mu}\underline{\boldsymbol{\sigma}}_{\boldsymbol{X}}^2\boldsymbol{\zeta}_{\mu}')^{-1}(\boldsymbol{\mu}^{view} - \boldsymbol{\zeta}_{\mu}\underline{\boldsymbol{\mu}}_{\boldsymbol{X}}), \tag{A.1.85}$$

and the updated expectation

$$\bar{\boldsymbol{\mu}}_{\boldsymbol{X}} = \bar{\boldsymbol{\mu}}_{\boldsymbol{X};\sigma} + \boldsymbol{\zeta}_{\mu}^{\dagger\prime}(\boldsymbol{\mu}^{view} - \boldsymbol{\zeta}_{\mu}\underline{\boldsymbol{\mu}}_{\boldsymbol{X}}), \tag{A.1.86}$$

according to Table 1.8, where here the pseudo-inverse $\boldsymbol{\zeta}_{\mu}^{\dagger}$ is under the base covariance $\underline{\boldsymbol{\sigma}}_{\boldsymbol{X}}^2$ according to (1.90).

A.1.11 Normal MRE update: views on expectations and covariances as projection

If we define the following $\bar{n} \times 1$ vector

$$\bar{\boldsymbol{\mu}}_{\boldsymbol{X};\sigma} \equiv \underline{\boldsymbol{\mu}}_{\boldsymbol{X}} + \boldsymbol{\zeta}_{\sigma}^{\dagger\prime}(\boldsymbol{\sigma}^{2view}(\boldsymbol{\zeta}_{\sigma}\underline{\boldsymbol{\sigma}}_{\boldsymbol{X}}^2\boldsymbol{\zeta}_{\sigma}')^{-1} - \mathbb{I}_{\bar{k}\times\bar{k}})\boldsymbol{\zeta}_{\sigma}\underline{\boldsymbol{\mu}}_{\boldsymbol{X}} \tag{A.1.87}$$

then, according to (1.57), we have

$$\bar{\boldsymbol{\mu}}_{\boldsymbol{X};\sigma} = \boldsymbol{\zeta}_{\sigma}^{\dagger\prime}\boldsymbol{\sigma}^{2view}(\boldsymbol{\zeta}_{\sigma}\underline{\boldsymbol{\sigma}}_{\boldsymbol{X}}^2\boldsymbol{\zeta}_{\sigma}')^{-1}\boldsymbol{\zeta}_{\sigma}\underline{\boldsymbol{\mu}}_{\boldsymbol{X}} + \mathfrak{P}_{\boldsymbol{\zeta}_{\sigma}}[\underline{\boldsymbol{\mu}}_{\boldsymbol{X}} - \boldsymbol{\zeta}_{\sigma}^{\dagger\prime}\boldsymbol{\sigma}^{2view}(\boldsymbol{\zeta}_{\sigma}\underline{\boldsymbol{\sigma}}_{\boldsymbol{X}}^2\boldsymbol{\zeta}_{\sigma}')^{-1}\boldsymbol{\zeta}_{\sigma}\underline{\boldsymbol{\mu}}_{\boldsymbol{X}}]. \tag{A.1.88}$$

Next, since we can rewrite the updated expectation (1.64) as follows

$$\bar{\boldsymbol{\mu}}_{\boldsymbol{X}} = \bar{\boldsymbol{\mu}}_{\boldsymbol{X};\sigma} + \bar{\boldsymbol{\zeta}}_{\mu}^{\dagger\prime}(\boldsymbol{\eta}_{\mu}^{view} - \boldsymbol{\zeta}_{\mu}\bar{\boldsymbol{\mu}}_{\boldsymbol{X};\sigma}), \tag{A.1.89}$$

which is consistent with the updating formula for the expectation (1.21) under $\boldsymbol{X} \sim N(\bar{\boldsymbol{\mu}}_{\boldsymbol{X};\sigma}, \bar{\boldsymbol{\sigma}}_{\boldsymbol{X}}^2)$ as new base distribution (1.6).

Hence (1.85) simply follows from (1.26).

Now, let us assume that the view variables are statistically independent under the base distribution as in (1.89).

Then, the pseudo inverses $\bar{\boldsymbol{\zeta}}_{\mu}^{\dagger}$ (1.65) and $\boldsymbol{\zeta}_{\mu}^{\dagger}$ (1.22) are the same (1.90) since from (1.66) we have

$$\boldsymbol{\zeta}_{\mu}\bar{\boldsymbol{\sigma}}_{\boldsymbol{X}}^2\boldsymbol{\zeta}_{\mu}' = \boldsymbol{\zeta}_{\mu}\underline{\boldsymbol{\sigma}}_{\boldsymbol{X}}^2\boldsymbol{\zeta}_{\mu}'. \tag{A.1.90}$$

Also, the projectors $\mathfrak{P}_{\boldsymbol{\zeta}_{\mu}}$ and $\mathfrak{P}_{\boldsymbol{\zeta}_{\sigma}}$ commutes (1.91), since we have

$$\begin{aligned} \mathfrak{P}_{\boldsymbol{\zeta}_{\mu}}[\mathfrak{P}_{\boldsymbol{\zeta}_{\sigma}}[\boldsymbol{x}]] &= (\mathbb{I}_{\bar{n}} - \boldsymbol{\zeta}_{\mu}^{\dagger\prime}\boldsymbol{\zeta}_{\mu})(\mathbb{I}_{\bar{n}} - \boldsymbol{\zeta}_{\sigma}^{\dagger\prime}\boldsymbol{\zeta}_{\sigma})\boldsymbol{x} \\ &= (\mathbb{I}_{\bar{n}} - \boldsymbol{\zeta}_{\mu}^{\dagger\prime}\boldsymbol{\zeta}_{\mu} - \boldsymbol{\zeta}_{\sigma}^{\dagger\prime}\boldsymbol{\zeta}_{\sigma})\boldsymbol{x} \\ &= \mathfrak{P}_{\boldsymbol{\zeta}_{\sigma}}[\mathfrak{P}_{\boldsymbol{\zeta}_{\mu}}[\boldsymbol{x}]], \end{aligned} \tag{A.1.91}$$

as follows because

$$\boldsymbol{\zeta}_{\mu}\boldsymbol{\zeta}_{\sigma}^{\dagger\prime} = \boldsymbol{\zeta}_{\mu}\underline{\boldsymbol{\sigma}}_{\boldsymbol{X}}^2\boldsymbol{\zeta}_{\sigma}'(\boldsymbol{\zeta}_{\sigma}\underline{\boldsymbol{\sigma}}_{\boldsymbol{X}}^2\boldsymbol{\zeta}_{\sigma}')^{-1} = \mathbf{0}. \tag{A.1.92}$$

Finally, the view variables are statistically independent also under the updated distribution (1.92), since from (1.66) we have

$$\begin{aligned} \boldsymbol{\zeta}_{\mu}\bar{\boldsymbol{\sigma}}_{\boldsymbol{X}}^2\boldsymbol{\zeta}_{\sigma}' &= \boldsymbol{\zeta}_{\mu}\underline{\boldsymbol{\sigma}}_{\boldsymbol{X}}^2\boldsymbol{\zeta}_{\sigma}' + \boldsymbol{\zeta}_{\mu}\boldsymbol{\zeta}_{\sigma}^{\dagger\prime}(\boldsymbol{\sigma}^{2view} - \boldsymbol{\zeta}_{\sigma}\underline{\boldsymbol{\sigma}}_{\boldsymbol{X}}^2\boldsymbol{\zeta}_{\sigma}')\boldsymbol{\zeta}_{\sigma}^{\dagger}\boldsymbol{\zeta}_{\sigma}' \\ &= \mathbf{0}, \end{aligned} \tag{A.1.93}$$

where the last row follows from the statistical independence under the base (1.89) and (A.1.92).

Appendix 2

Here we discuss some technical results of Chapter 2.

A.2.1 Dual Lagrangian: gradient and Hessian

According to the arguments in [A.1.3] the dual Lagrangian $\mathcal{L}(\boldsymbol{t};\boldsymbol{\eta}^{view})$ (2.7) is a convex function which we can write in terms of the yet-to-be defined expectation (A.1.22)

$$\begin{aligned}\boldsymbol{\mu}_{\boldsymbol{X}} &\equiv \mu(\boldsymbol{t}) = \boldsymbol{\sigma}^2_{\boldsymbol{X}}(\underline{\boldsymbol{\theta}}_{\boldsymbol{X};\mu} + \boldsymbol{\zeta}'_{\mu}\boldsymbol{t}_{\mu}) \\ &= \underline{\boldsymbol{\mu}}_{\boldsymbol{X}} + \boldsymbol{p}_{\sigma}\underline{\boldsymbol{\mu}}_{\boldsymbol{X}} + \boldsymbol{\sigma}^2_{\boldsymbol{X}}\boldsymbol{\zeta}'_{\mu}\boldsymbol{t}_{\mu},\end{aligned} \tag{A.2.1}$$

and the yet-to-be defined covariance (A.1.20)

$$\boldsymbol{\sigma}^2_{\boldsymbol{X}} \equiv \sigma^2(\boldsymbol{t}) = \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}} + \boldsymbol{p}_{\sigma}\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}, \tag{A.2.2}$$

where $\boldsymbol{p}_{\sigma}$ is the following $\bar{n}\times\bar{n}$ matrix (A.1.21)

$$\boldsymbol{p}_{\sigma} \equiv \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}\boldsymbol{\zeta}'_{\sigma}(\frac{1}{2}(\boldsymbol{t}_{\sigma,\sigma})^{-1} - \boldsymbol{\zeta}_{\sigma}\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}\boldsymbol{\zeta}'_{\sigma})^{-1}\boldsymbol{\zeta}_{\sigma}. \tag{A.2.3}$$

According to the above, note that $\boldsymbol{\mu}_{\boldsymbol{X}}$ depends on *both* $\boldsymbol{t}_{\mu}$ and $\boldsymbol{t}_{\sigma,\sigma}$, while $\boldsymbol{\sigma}^2_{\boldsymbol{X}}$ depends only on $\boldsymbol{t}_{\sigma,\sigma}$.

Let us focus on the first derivatives.

We recall that the view variables in this context reads

$$\zeta^{view}(\boldsymbol{X}) \equiv \begin{pmatrix} \boldsymbol{\zeta}_{\mu}\boldsymbol{X} \\ \boldsymbol{\zeta}_{\sigma,\sigma}vec(\boldsymbol{X}\boldsymbol{X}') \end{pmatrix}, \tag{A.2.4}$$

and, according to (29), the gradient of the dual Lagrangian is the expectation of the shifted view variables $\zeta^{view}(\boldsymbol{X}) - \boldsymbol{\eta}^{view}$, under the yet-to-be-defined distribution $f_{\boldsymbol{X}} = f_{\boldsymbol{t}}$ (24).

This means in particular that the gradient with respect to $\boldsymbol{t}_{\mu}$ must read

$$\begin{aligned}\nabla_{\boldsymbol{t}_{\mu}}\mathcal{L}(\boldsymbol{t};\boldsymbol{\eta}^{view}) &= \mathbb{E}^{f_{\boldsymbol{X}}}\{\boldsymbol{\zeta}_{\mu}\boldsymbol{X} - \boldsymbol{\eta}^{view}_{\mu}\} \\ &= \boldsymbol{\zeta}_{\mu}\boldsymbol{\mu}_{\boldsymbol{X}} - \boldsymbol{\eta}^{view}_{\mu};\end{aligned} \tag{A.2.5}$$

and similar the gradient with respect to $\boldsymbol{t}_{\sigma,\sigma}$ must read

$$\begin{aligned}\nabla_{\boldsymbol{t}_{\sigma,\sigma}}\mathcal{L}(\boldsymbol{t};\boldsymbol{\eta}^{view}) &= \mathbb{E}^{f_{\boldsymbol{X}}}\{\boldsymbol{\zeta}_{\sigma,\sigma}vec(\boldsymbol{X}\boldsymbol{X}') - vec(\boldsymbol{\eta}^{view}_{\sigma,\sigma})\} \\ &= \boldsymbol{\zeta}_{\sigma,\sigma}vec(\mathbb{E}^{f_{\boldsymbol{X}}}\{\boldsymbol{X}\boldsymbol{X}'\}) - vec(\boldsymbol{\eta}^{view}_{\sigma,\sigma}) \\ &= \boldsymbol{\zeta}_{\sigma,\sigma}vec(\boldsymbol{\sigma}^2_{\boldsymbol{X}} + \boldsymbol{\mu}_{\boldsymbol{X}}\boldsymbol{\mu}'_{\boldsymbol{X}}) - vec(\boldsymbol{\eta}^{view}_{\sigma,\sigma}) \\ &= vec(\boldsymbol{\zeta}_{\sigma}(\boldsymbol{\sigma}^2_{\boldsymbol{X}} + \boldsymbol{\mu}_{\boldsymbol{X}}\boldsymbol{\mu}'_{\boldsymbol{X}})\boldsymbol{\zeta}'_{\sigma} - \boldsymbol{\eta}^{view}_{\sigma,\sigma}),\end{aligned} \tag{A.2.6}$$

where we used the definition of the $\bar{k}^2_\sigma \times \bar{n}^2$ matrix $\boldsymbol{\zeta}_{\sigma,\sigma}$ (1.8) and the property of the Kronecker product $\otimes$ with respect to the vectorization operator vec, or

$$vec(\boldsymbol{x}\boldsymbol{y}\boldsymbol{z}) = (\boldsymbol{z}' \otimes \boldsymbol{x})vec(\boldsymbol{y}), \tag{A.2.7}$$

for any conformable matrices $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{z}$. Refer to [Magnus and Neudecker, 1979] for details.

Let us focus on the second derivatives.

We recall, according to (30), that the Hessian of the dual Lagrangian is the covariance of the view variables $\zeta^{view}(\boldsymbol{X})$, under the yet-to-be-defined distribution $f_{\boldsymbol{X}} = f_{\boldsymbol{t}}$ (24)

$$\boldsymbol{X} \sim f_{\boldsymbol{X}} \Leftrightarrow N(\boldsymbol{\mu}_{\boldsymbol{X}}, \boldsymbol{\sigma}^2_{\boldsymbol{X}}). \tag{A.2.8}$$

Then the Hessian with respect to $\boldsymbol{t}_\mu$ reads

$$\begin{aligned}\nabla^2_{\boldsymbol{t}_\mu,\boldsymbol{t}_\mu}\mathcal{L}(\boldsymbol{t};\boldsymbol{\eta}^{view}) &= \mathbb{C}v^{f_{\boldsymbol{X}}}\{\boldsymbol{\zeta}_\mu\boldsymbol{X}\} \\ &= \boldsymbol{\zeta}_\mu\boldsymbol{\sigma}^2_{\boldsymbol{X}}\boldsymbol{\zeta}'_\mu.\end{aligned} \tag{A.2.9}$$

In order to compute Hessian with respect to the cross-variables $(\boldsymbol{t}_{\sigma,\sigma}, \boldsymbol{t}_\mu)$, we start noting from (A.2.6) that the sequential differential of the dual Lagrangian $\mathcal{L}(\boldsymbol{t};\boldsymbol{\eta}^{view})$ (2.7) reads

$$\begin{aligned}d_{\boldsymbol{t}_\mu}(d\boldsymbol{t}_{\sigma,\sigma}\mathcal{L}(\boldsymbol{t};\boldsymbol{\eta}^{view})) &= d_{\boldsymbol{t}_\mu}vec((\boldsymbol{\zeta}_\sigma(\boldsymbol{\sigma}^2_{\boldsymbol{X}} + \boldsymbol{\mu}_{\boldsymbol{X}}\boldsymbol{\mu}'_{\boldsymbol{X}})\boldsymbol{\zeta}'_\sigma - \boldsymbol{\eta}^{view}_{\sigma,\sigma}))'vec(d\boldsymbol{t}_{\sigma,\sigma}) \\ &= vec(\boldsymbol{\zeta}_\sigma d_{\boldsymbol{t}_\mu}(\boldsymbol{\mu}_{\boldsymbol{X}}\boldsymbol{\mu}'_{\boldsymbol{X}})\boldsymbol{\zeta}'_\sigma)'vec(d\boldsymbol{t}_{\sigma,\sigma}) \\ &= vec(\boldsymbol{\zeta}_\sigma(d_{\boldsymbol{t}_\mu}\boldsymbol{\mu}_{\boldsymbol{X}}\boldsymbol{\mu}'_{\boldsymbol{X}} + \boldsymbol{\mu}_{\boldsymbol{X}}d_{\boldsymbol{t}_\mu}\boldsymbol{\mu}'_{\boldsymbol{X}})\boldsymbol{\zeta}'_\sigma)'vec(d\boldsymbol{t}_{\sigma,\sigma}),\end{aligned} \tag{A.2.10}$$

where the second row follows because $\boldsymbol{\sigma}^2_{\boldsymbol{X}}$ does *not* depend only on $\boldsymbol{t}_\mu$.

Now, since the first order differential of yet-to-be defined expectation (A.2.1) with respect to $\boldsymbol{t}_\mu$ reads

$$d_{\boldsymbol{t}_\mu}\boldsymbol{\mu}_{\boldsymbol{X}} = \boldsymbol{\sigma}^2_{\boldsymbol{X}}\boldsymbol{\zeta}'_\mu d\boldsymbol{t}_\mu, \tag{A.2.11}$$

we derive

$$d_{\boldsymbol{t}_\mu}\boldsymbol{\mu}_{\boldsymbol{X}}\boldsymbol{\mu}'_{\boldsymbol{X}} = \boldsymbol{\sigma}^2_{\boldsymbol{X}}\boldsymbol{\zeta}'_\mu d\boldsymbol{t}_\mu\boldsymbol{\mu}'_{\boldsymbol{X}}, \tag{A.2.12}$$

and

$$\boldsymbol{\mu}_{\boldsymbol{X}}d_{\boldsymbol{t}_\mu}\boldsymbol{\mu}'_{\boldsymbol{X}} = \boldsymbol{\mu}_{\boldsymbol{X}}d\boldsymbol{t}'_\mu\boldsymbol{\zeta}_\mu\boldsymbol{\sigma}^2_{\boldsymbol{X}}. \tag{A.2.13}$$

This means that

$$\begin{aligned} vec(\boldsymbol{\zeta}_\sigma(d_{\boldsymbol{t}_\mu}\boldsymbol{\mu}_{\boldsymbol{X}}\boldsymbol{\mu}'_{\boldsymbol{X}} + \boldsymbol{\mu}_{\boldsymbol{X}} d_{\boldsymbol{t}_\mu}\boldsymbol{\mu}'_{\boldsymbol{X}})\boldsymbol{\zeta}'_\sigma) &= vec(\boldsymbol{\zeta}_\sigma(\boldsymbol{\sigma}^2_{\boldsymbol{X}}\boldsymbol{\zeta}'_\mu d\boldsymbol{t}_\mu\boldsymbol{\mu}'_{\boldsymbol{X}} + \boldsymbol{\mu}_{\boldsymbol{X}} d\boldsymbol{t}'_\mu\boldsymbol{\zeta}_\mu\boldsymbol{\sigma}^2_{\boldsymbol{X}})\boldsymbol{\zeta}'_\sigma) \quad \text{(A.2.14)} \\ &= ((\boldsymbol{\zeta}_\sigma\boldsymbol{\mu}_{\boldsymbol{X}} \otimes \boldsymbol{\zeta}_\sigma\boldsymbol{\sigma}^2_{\boldsymbol{X}}\boldsymbol{\zeta}'_\mu) + (\boldsymbol{\zeta}_\sigma\boldsymbol{\sigma}^2_{\boldsymbol{X}}\boldsymbol{\zeta}'_\mu \otimes \boldsymbol{\zeta}_\sigma\boldsymbol{\mu}_{\boldsymbol{X}}))d\boldsymbol{t}_\mu \\ &= (\boldsymbol{\zeta}_{\sigma,\sigma}(\boldsymbol{\mu}_{\boldsymbol{X}} \otimes \boldsymbol{\sigma}^2_{\boldsymbol{X}}\boldsymbol{\zeta}'_\mu) + (\boldsymbol{\zeta}_{\sigma,\sigma}(\boldsymbol{\sigma}^2_{\boldsymbol{X}}\boldsymbol{\zeta}'_\mu \otimes \boldsymbol{\mu}_{\boldsymbol{X}}))d\boldsymbol{t}_\mu \\ &= \boldsymbol{\zeta}_{\sigma,\sigma}[(\boldsymbol{\mu}_{\boldsymbol{X}} \otimes \boldsymbol{\sigma}^2_{\boldsymbol{X}}\boldsymbol{\zeta}'_\mu) + (\boldsymbol{\sigma}^2_{\boldsymbol{X}}\boldsymbol{\zeta}'_\mu \otimes \boldsymbol{\mu}_{\boldsymbol{X}})]d\boldsymbol{t}_\mu \\ &= \boldsymbol{\zeta}_{\sigma,\sigma}[(\boldsymbol{\mu}_{\boldsymbol{X}} \otimes \boldsymbol{\sigma}^2_{\boldsymbol{X}}) + (\boldsymbol{\sigma}^2_{\boldsymbol{X}} \otimes \boldsymbol{\mu}_{\boldsymbol{X}})]\boldsymbol{\zeta}'_\mu d\boldsymbol{t}_\mu, \end{aligned}$$

where in the second row we applied (A.2.7), or

$$vec(\boldsymbol{\zeta}_\sigma\boldsymbol{\sigma}^2_{\boldsymbol{X}}\boldsymbol{\zeta}'_\mu d\boldsymbol{t}_\mu\boldsymbol{\mu}'_{\boldsymbol{X}}\boldsymbol{\zeta}'_\sigma) = (\boldsymbol{\zeta}_\sigma\boldsymbol{\mu}_{\boldsymbol{X}} \otimes \boldsymbol{\zeta}_\sigma\boldsymbol{\sigma}^2_{\boldsymbol{X}}\boldsymbol{\zeta}'_\mu)d\boldsymbol{t}_\mu, \tag{A.2.15}$$

and

$$vec(\boldsymbol{\zeta}_\sigma\boldsymbol{\mu}_{\boldsymbol{X}} d\boldsymbol{t}'_\mu\boldsymbol{\zeta}_\mu\boldsymbol{\sigma}^2_{\boldsymbol{X}}\boldsymbol{\zeta}'_\sigma) = (\boldsymbol{\zeta}_\sigma\boldsymbol{\sigma}^2_{\boldsymbol{X}}\boldsymbol{\zeta}'_\mu \otimes \boldsymbol{\zeta}_\sigma\boldsymbol{\mu}_{\boldsymbol{X}})d\boldsymbol{t}_\mu, \tag{A.2.16}$$

note that $vec(d\boldsymbol{t}_\mu) = vec(d\boldsymbol{t}'_\mu) = d\boldsymbol{t}_\mu$; and in the third and fifth row we used the property of the Kronecker product $\otimes$ with respect to standard matrices product

$$(\boldsymbol{x} \otimes \boldsymbol{y})(\boldsymbol{w} \otimes \boldsymbol{z}) = (\boldsymbol{x}\boldsymbol{w}) \otimes (\boldsymbol{y}\boldsymbol{z}), \tag{A.2.17}$$

for any conformable matrices $\boldsymbol{x}$, $\boldsymbol{y}$, $\boldsymbol{w}$ and $\boldsymbol{z}$. Refer to [Magnus and Neudecker, 1979] for details.

To conclude, since

$$d_{\boldsymbol{t}_\mu}(d\boldsymbol{t}_{\sigma,\sigma}\mathcal{L}(\boldsymbol{t};\boldsymbol{\eta}^{view})) = d\boldsymbol{t}'_\mu(\nabla^2_{\boldsymbol{t}_{\sigma,\sigma},\boldsymbol{t}_\mu}\mathcal{L}(\boldsymbol{t};\boldsymbol{\eta}^{view}))' vec(d\boldsymbol{t}_{\sigma,\sigma}), \tag{A.2.18}$$

then we must have

$$\nabla^2_{\boldsymbol{t}_{\sigma,\sigma},\boldsymbol{t}_\mu}\mathcal{L}(\boldsymbol{t};\boldsymbol{\eta}^{view}) = \boldsymbol{\zeta}_{\sigma,\sigma}[(\boldsymbol{\mu}_{\boldsymbol{X}} \otimes \boldsymbol{\sigma}^2_{\boldsymbol{X}}) + (\boldsymbol{\sigma}^2_{\boldsymbol{X}} \otimes \boldsymbol{\mu}_{\boldsymbol{X}})]\boldsymbol{\zeta}'_\mu \tag{A.2.19}$$

This structure is consistent with what we expect. Indeed from (30) the Hessian $\nabla^2_{\boldsymbol{t}_{\sigma,\sigma},\boldsymbol{t}_\mu}\mathcal{L}(\boldsymbol{t};\boldsymbol{\eta}^{view})$ is the cross-covariance the view variables $\boldsymbol{\zeta}_{\sigma,\sigma} vec(\boldsymbol{X}\boldsymbol{X}')$ and $\boldsymbol{\zeta}_\mu\boldsymbol{X}$

$$\begin{aligned} \nabla^2_{\boldsymbol{t}_{\sigma,\sigma},\boldsymbol{t}_\mu}\mathcal{L}(\boldsymbol{t};\boldsymbol{\eta}^{view}) &= \mathbb{C}v^{f\boldsymbol{X}}\{\boldsymbol{\zeta}_{\sigma,\sigma} vec(\boldsymbol{X}\boldsymbol{X}'), \boldsymbol{\zeta}_\mu\boldsymbol{X}\}\boldsymbol{\zeta}'_\mu \\ &= \boldsymbol{\zeta}_{\sigma,\sigma}\mathbb{C}v^{f\boldsymbol{X}}\{vec(\boldsymbol{X}\boldsymbol{X}'), \boldsymbol{X}\}\boldsymbol{\zeta}'_\mu. \end{aligned} \tag{A.2.20}$$

Finally, according to (30), the Hessian with respect to $\boldsymbol{t}_{\sigma,\sigma}$ reads

$$\begin{aligned} \nabla^2_{\boldsymbol{t}_{\sigma,\sigma},\boldsymbol{t}_{\sigma,\sigma}}\mathcal{L}(\boldsymbol{t};\boldsymbol{\eta}^{view}) &= \mathbb{C}v^{f\boldsymbol{X}}\{\boldsymbol{\zeta}_{\sigma,\sigma} vec(\boldsymbol{X}\boldsymbol{X}')\} \\ &= \boldsymbol{\zeta}_{\sigma,\sigma}\mathbb{C}v^{f\boldsymbol{X}}\{vec(\boldsymbol{X}\boldsymbol{X}')\}\boldsymbol{\zeta}_{\sigma,\sigma}. \end{aligned} \tag{A.2.21}$$

In order to develop the covariance in (A.2.21), we define the following $\bar{n} \times \bar{n}$ matrix-variate variable

$$\boldsymbol{W}^2 \equiv (\boldsymbol{X} - \boldsymbol{\mu}_{\boldsymbol{X}})(\boldsymbol{X} - \boldsymbol{\mu}_{\boldsymbol{X}})', \tag{A.2.22}$$

which, under the yet-to-be-defined distribution $f_{\boldsymbol{X}} = f_{\boldsymbol{t}}$ (24), follows by construction a Wishart distribution with 1 degree of freedom and dispersion parameter $\boldsymbol{\sigma}^2_{\boldsymbol{X}}$

$$\boldsymbol{W}^2 \sim Wishart(1, \boldsymbol{\sigma}^2_{\boldsymbol{X}}). \tag{A.2.23}$$

Refer also to [Anderson, 1984] for more details.

From (A.2.22), we can decompose the variable $\boldsymbol{W}^2$ as the sum of the following terms

$$\boldsymbol{W}^2 = \boldsymbol{X}\boldsymbol{X}' - \boldsymbol{\mu}_{\boldsymbol{X}}\boldsymbol{X}' - \boldsymbol{X}\boldsymbol{\mu}'_{\boldsymbol{X}} + \boldsymbol{\mu}_{\boldsymbol{X}}\boldsymbol{\mu}'_{\boldsymbol{X}}, \tag{A.2.24}$$

from which we recover

$$\begin{aligned} \mathbb{C}v^{f_{\boldsymbol{X}}}\left\{vec(\boldsymbol{W}^2)\right\} &= \mathbb{C}v^{f_{\boldsymbol{X}}}\left\{vec(\boldsymbol{X}\boldsymbol{X}' - \boldsymbol{\mu}_{\boldsymbol{X}}\boldsymbol{X}' - \boldsymbol{X}\boldsymbol{\mu}'_{\boldsymbol{X}} + \boldsymbol{\mu}_{\boldsymbol{X}}\boldsymbol{\mu}'_{\boldsymbol{X}})\right\} \\ &= \mathbb{C}v^{f_{\boldsymbol{X}}}\left\{vec(\boldsymbol{X}\boldsymbol{X}') - vec(\boldsymbol{\mu}_{\boldsymbol{X}}\boldsymbol{X}' - \boldsymbol{X}\boldsymbol{\mu}'_{\boldsymbol{X}})\right\} \\ &= \mathbb{C}v^{f_{\boldsymbol{X}}}\left\{vec(\boldsymbol{X}\boldsymbol{X}')\right\} + \boldsymbol{\alpha}^2 - \boldsymbol{\beta} - \boldsymbol{\beta}'; \end{aligned} \tag{A.2.25}$$

where we defined

$$\boldsymbol{\alpha}^2 \equiv \mathbb{C}v^{f_{\boldsymbol{X}}}\left\{vec(\boldsymbol{\mu}_{\boldsymbol{X}}\boldsymbol{X}' + \boldsymbol{X}\boldsymbol{\mu}'_{\boldsymbol{X}})\right\} \quad \boldsymbol{\beta} \equiv \mathbb{C}v^{f_{\boldsymbol{X}}}\left\{vec(\boldsymbol{X}\boldsymbol{X}'), vec(\boldsymbol{\mu}_{\boldsymbol{X}}\boldsymbol{X}' + \boldsymbol{X}\boldsymbol{\mu}'_{\boldsymbol{X}})\right\}; \tag{A.2.26}$$

and where we used the bilinearity of the covariance operator

$$\mathbb{C}v\{\boldsymbol{X} + \boldsymbol{Y}\} = \mathbb{C}v\{\boldsymbol{X}\} + \mathbb{C}v\{\boldsymbol{X}, \boldsymbol{Y}\} + \mathbb{C}v\{\boldsymbol{X}, \boldsymbol{Y}\}' + \mathbb{C}v\{\boldsymbol{Y}\}. \tag{A.2.27}$$

Hence the desired covariance in (A.2.21) reads

$$\mathbb{C}v^{f_{\boldsymbol{X}}}\left\{vec(\boldsymbol{X}\boldsymbol{X}')\right\} = \mathbb{C}v^{f_{\boldsymbol{X}}}\left\{vec(\boldsymbol{W}^2)\right\} + \boldsymbol{\beta} + \boldsymbol{\beta}' - \boldsymbol{\alpha}^2. \tag{A.2.28}$$

Now the covariance of the Wishart distribution (A.2.23) reads

$$\mathbb{C}v^{f_{\boldsymbol{X}}}\left\{vec(\boldsymbol{W}^2)\right\} = (\mathbb{I}_{\bar{n}^2} + \mathbb{K}_{\bar{n},\bar{n}})(\boldsymbol{\sigma}^2_{\boldsymbol{X}} \otimes \boldsymbol{\sigma}^2_{\boldsymbol{X}}), \tag{A.2.29}$$

where $\mathbb{K}_{\bar{n},\bar{n}}$ denotes the $\bar{n}^2 \times \bar{n}^2$ commutation matrix [Magnus and Neudecker, 1979]. Refer also to [Anderson, 1984] for more details.

From the other hand, the term $\boldsymbol{\beta}$ in (A.2.26) explicitly reads

$$\begin{aligned} \boldsymbol{\beta} &= \mathbb{C}v^{f_{\boldsymbol{X}}}\left\{vec(\boldsymbol{X}\boldsymbol{X}'), vec(\boldsymbol{\mu}_{\boldsymbol{X}}\boldsymbol{X}')\right\} + \mathbb{C}v^{f_{\boldsymbol{X}}}\left\{vec(\boldsymbol{X}\boldsymbol{X}'), vec(\boldsymbol{X}\boldsymbol{\mu}'_{\boldsymbol{X}})\right\} \\ &= \mathbb{C}v^{f_{\boldsymbol{X}}}\left\{vec(\boldsymbol{X}\boldsymbol{X}'), (\mathbb{I}_n \otimes \boldsymbol{\mu}_{\boldsymbol{X}})\boldsymbol{X}\right\} + \mathbb{C}v^{f_{\boldsymbol{X}}}\left\{vec(\boldsymbol{X}\boldsymbol{X}'), (\boldsymbol{\mu}_{\boldsymbol{X}} \otimes \mathbb{I}_{\bar{n}})\boldsymbol{X}\right\} \\ &= \mathbb{C}v^{f_{\boldsymbol{X}}}\left\{vec(\boldsymbol{X}\boldsymbol{X}'), \boldsymbol{X}\right\}(\mathbb{I}_{\bar{n}} \otimes \boldsymbol{\mu}_{\boldsymbol{X}})' + \mathbb{C}v^{f_{\boldsymbol{X}}}\left\{vec(\boldsymbol{X}\boldsymbol{X}'), \boldsymbol{X}\right\}(\boldsymbol{\mu}_{\boldsymbol{X}} \otimes \mathbb{I}_{\bar{n}})' \\ &= [(\boldsymbol{\mu}_{\boldsymbol{X}} \otimes \boldsymbol{\sigma}^2_{\boldsymbol{X}}) + (\boldsymbol{\sigma}^2_{\boldsymbol{X}} \otimes \boldsymbol{\mu}_{\boldsymbol{X}})][(\mathbb{I}_{\bar{n}} \otimes \boldsymbol{\mu}'_{\boldsymbol{X}}) + (\boldsymbol{\mu}'_{\boldsymbol{X}} \otimes \mathbb{I}_{\bar{n}})], \end{aligned} \tag{A.2.30}$$

where in the second row we used (*A.2.7*), i.e.

$$\begin{aligned} vec(\boldsymbol{\mu}_{\boldsymbol{X}}\boldsymbol{X}') &= (\mathbb{I}_n \otimes \boldsymbol{\mu}_{\boldsymbol{X}})vec(\boldsymbol{X}') \\ &= (\mathbb{I}_n \otimes \boldsymbol{\mu}_{\boldsymbol{X}})\boldsymbol{X}, \end{aligned} \tag{A.2.31}$$

and

$$\begin{aligned} vec(\boldsymbol{X}\boldsymbol{\mu}'_{\boldsymbol{X}}) &= (\boldsymbol{\mu}_{\boldsymbol{X}} \otimes \mathbb{I}_{\bar{n}})vec(\boldsymbol{X}) \\ &= (\boldsymbol{\mu}_{\boldsymbol{X}} \otimes \mathbb{I}_{\bar{n}})\boldsymbol{X}, \end{aligned} \tag{A.2.32}$$

and the expression of the cross-covariance $\mathbb{C}v^{f_{\boldsymbol{X}}}\{vec(\boldsymbol{X}\boldsymbol{X}'), \boldsymbol{X}\} = (\boldsymbol{\mu}_{\boldsymbol{X}} \otimes \boldsymbol{\sigma}^2_{\boldsymbol{X}}) + (\boldsymbol{\sigma}^2_{\boldsymbol{X}} \otimes \boldsymbol{\mu}_{\boldsymbol{X}})$ as follows from (A.2.19)-(A.2.20); and in the last row we used the property of the Kronecker product $\otimes$ with respect to transpositions

$$(\boldsymbol{x} \otimes \boldsymbol{y})' = \boldsymbol{x}' \otimes \boldsymbol{y}'. \tag{A.2.33}$$

Finally the term $\boldsymbol{\alpha}^2$ in (A.2.26) becomes

$$\begin{aligned} \boldsymbol{\alpha}^2 &= \mathbb{C}v^{f_{\boldsymbol{X}}}\{vec(\boldsymbol{\mu}_{\boldsymbol{X}}\boldsymbol{X}') + vec(\boldsymbol{X}\boldsymbol{\mu}'_{\boldsymbol{X}})\} && \text{(A.2.34)} \\ &= \mathbb{C}v^{f_{\boldsymbol{X}}}\{(\mathbb{I}_n \otimes \boldsymbol{\mu}_{\boldsymbol{X}})\boldsymbol{X} + (\boldsymbol{\mu}_{\boldsymbol{X}} \otimes \mathbb{I}_{\bar{n}})\boldsymbol{X}\} \\ &= \mathbb{C}v^{f_{\boldsymbol{X}}}\{[(\mathbb{I}_n \otimes \boldsymbol{\mu}_{\boldsymbol{X}}) + (\boldsymbol{\mu}_{\boldsymbol{X}} \otimes \mathbb{I}_{\bar{n}})]\boldsymbol{X}\} \\ &= [(\mathbb{I}_n \otimes \boldsymbol{\mu}_{\boldsymbol{X}}) + (\boldsymbol{\mu}_{\boldsymbol{X}} \otimes \mathbb{I}_{\bar{n}})]\boldsymbol{\sigma}^2_{\boldsymbol{X}}[(\mathbb{I}_n \otimes \boldsymbol{\mu}'_{\boldsymbol{X}}) + (\boldsymbol{\mu}'_{\boldsymbol{X}} \otimes \mathbb{I}_{\bar{n}})], && \text{(A.2.35)} \end{aligned}$$

as follows from (A.2.31)-(A.2.32) and (A.2.33).

To conclude, from the above results it is easy to verify that the Hessian reads as in (2.23), where the multivariate function $\epsilon(\boldsymbol{t})$ is

$$\epsilon(\boldsymbol{t}) \equiv \boldsymbol{\beta} + \boldsymbol{\beta}' - \boldsymbol{\alpha}^2. \tag{A.2.36}$$

A.2.2 Inequality views on expectations

Under normality of the base distribution (2.1) and inequality views of expectations as in (2.9)

$$f_{\boldsymbol{X}} \in \mathcal{C}_{\boldsymbol{X}}: \qquad \mathbb{E}^{f_{\boldsymbol{X}}}\{\boldsymbol{\zeta}_\mu \boldsymbol{X}\} \leq \boldsymbol{\eta}^{view}_\mu, \tag{A.2.37}$$

the yet-to-be-defined MRE updated distribution (23) must be normal in turn [A.1.1]

$$\boldsymbol{X} \sim N(\boldsymbol{\mu}_{\boldsymbol{X}}, \boldsymbol{\sigma}^2_{\boldsymbol{X}}), \tag{A.2.38}$$

with the yet-to-be-defined expectation (A.1.22)

$$\boldsymbol{\mu}_{\boldsymbol{X}} = \underline{\boldsymbol{\mu}}_{\boldsymbol{X}} + \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}\boldsymbol{\zeta}'_\mu \boldsymbol{t}_\mu, \tag{A.2.39}$$

and yet-to-be-defined covariance (A.1.20)

$$\boldsymbol{\sigma}^2_{\boldsymbol{X}} = \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}. \tag{A.2.40}$$

In particular, the optimal Lagrange multipliers $\boldsymbol{\theta}^{view}_\mu$ are defined as follows (27)

$$\boldsymbol{\theta}^{view} \equiv \underset{\boldsymbol{t}_\mu \leqq \boldsymbol{0}}{\operatorname{argmin}} \mathcal{L}(\boldsymbol{t}_\mu; \boldsymbol{\eta}^{view}_\mu), \tag{A.2.41}$$

where the dual Lagrangian (28), which here reads as in (A.1.15), simplifies to

$$\begin{aligned}\mathcal{L}(\boldsymbol{t}_\mu;\boldsymbol{\eta}^{view}) &= \frac{1}{2}[\boldsymbol{\mu}'_{\boldsymbol{X}}(\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}})^{-1}\boldsymbol{\mu}_{\boldsymbol{X}} - \underline{\boldsymbol{\mu}}'_{\boldsymbol{X}}(\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}})^{-1}\underline{\boldsymbol{\mu}}_{\boldsymbol{X}}] - \boldsymbol{t}'_\mu\boldsymbol{\eta}^{view}_\mu \\ &= \frac{1}{2}(\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}\boldsymbol{\zeta}'_\mu\boldsymbol{t}_\mu)'(\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}})^{-1}\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}\boldsymbol{\zeta}'_\mu\boldsymbol{t}_\mu + \boldsymbol{t}'_\mu\boldsymbol{\zeta}_\mu\underline{\boldsymbol{\mu}}_{\boldsymbol{X}} - \boldsymbol{t}'_\mu\boldsymbol{\eta}^{view}_\mu \\ &= \frac{1}{2}\boldsymbol{t}'_\mu(\boldsymbol{\zeta}_\mu\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}\boldsymbol{\zeta}'_\mu)\boldsymbol{t}_\mu + \boldsymbol{t}'_\mu(\boldsymbol{\zeta}_\mu\underline{\boldsymbol{\mu}}_{\boldsymbol{X}} - \boldsymbol{\eta}^{view}_\mu),\end{aligned} \tag{A.2.42}$$

where in the second row we used (A.2.39).

A.2.3 Inequality views on expectations as mean-variance problem

By definition of factor portfolios $\boldsymbol{Z} \equiv \boldsymbol{\zeta}_\mu\boldsymbol{X} - \boldsymbol{\eta}^{view}_\mu$ (2.13) and affine equivariance of the expectation operator, we obtain

$$\underline{\mathbb{E}}\{\boldsymbol{Z}\} = \boldsymbol{\zeta}_\mu\underline{\mathbb{E}}\{\boldsymbol{X}\} - \boldsymbol{\eta}^{view}_\mu = \boldsymbol{\zeta}_\mu\underline{\boldsymbol{\mu}}_{\boldsymbol{X}} - \boldsymbol{\eta}^{view}_\mu. \tag{A.2.43}$$

Similar to the above, using the affine equivariance of the covariance operator, we obtain

$$\underline{\mathbb{C}v}\{\boldsymbol{Z}\} = \boldsymbol{\zeta}_\mu\underline{\mathbb{C}v}\{\boldsymbol{X}\}\boldsymbol{\zeta}'_\mu = \boldsymbol{\zeta}_\mu\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}\boldsymbol{\zeta}'_\mu. \tag{A.2.44}$$

Then changing the optimizing variable as follows

$$\boldsymbol{h} \equiv -\boldsymbol{t}_\mu, \tag{A.2.45}$$

we can rewrite the dual Lagrangian (2.12) as follows

$$\begin{aligned}\mathcal{L}(\boldsymbol{t}_\mu;\boldsymbol{\eta}^{view}_\mu) &= \frac{1}{2}\boldsymbol{h}'(\boldsymbol{\zeta}_\mu\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}\boldsymbol{\zeta}'_\mu)\boldsymbol{h} - \boldsymbol{h}'(\boldsymbol{\zeta}_\mu\underline{\boldsymbol{\mu}}_{\boldsymbol{X}} - \boldsymbol{\eta}^{view}_\mu) \\ &= \frac{1}{2}\underline{\mathbb{V}}\{Z_{\boldsymbol{h}}\} - \underline{\mathbb{E}}\{Z_{\boldsymbol{h}}\},\end{aligned} \tag{A.2.46}$$

where in the last row we applied again the affine equivariance to the variable $Z_{\boldsymbol{h}} \equiv \boldsymbol{h}'\boldsymbol{Z}$.

Finally, since minimizing the dual Lagrangian as in (2.6) is equivalent to maximizing its opposite, we easily obtain the desired result (2.14).

A.2.4 Ranking views as inequality views on expectations

Under the restriction $\mathbb{C}v^{f_{\boldsymbol{X}}}\{\boldsymbol{X}\} = \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}$, the ranking views (2.38) are equivalent to the following inequality statements

$$\frac{\mathbb{E}^{f_{\boldsymbol{X}}}\{X_n\}}{[\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X};vol}]_n} - \frac{\mathbb{E}^{f_{\boldsymbol{X}}}\{X_{n+1}\}}{[\underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X};vol}]_{n+1}} \leq 0, \tag{A.2.47}$$

for any $n = 1, \ldots, \bar{n} - 1$, which are clearly equivalent to the inequality views of expectations (2.9) for $\boldsymbol{\eta}_{\mu}^{view} = \mathbf{0}$ (and strict inequalities for $\boldsymbol{\eta}_{\mu}^{view} < \mathbf{0}$).

Moreover, since the condition $\mathbb{C}v^{f_{\boldsymbol{X}}}\{\boldsymbol{X}\} = \underline{\boldsymbol{\sigma}}^2_{\boldsymbol{X}}$ is already satisfied by the MRE solution under the only inequality views of expectations (2.9) [A.2.2], this implies the equivalence between (2.47) and (2.9).

Appendix 3

Here we discuss some technical results of Chapter 3.

A.3.1 Log-partition function: equivalent formulation

From the law of the unconscious statistician the following properties holds

$$\mathbb{E}\{\zeta^{view}(\boldsymbol{X})\} = \int_{\mathbb{R}^{\bar{n}}} \zeta^{view}(\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{x}, \tag{A.3.1}$$

for any arbitrary distribution $\boldsymbol{X} \sim f_{\boldsymbol{X}}$ and transformation ζ^{view}.

Hence the log-partition function (25) can be re-written as follows

$$\psi(\boldsymbol{t}) \equiv \ln \int_{\mathbb{R}^{\bar{n}}} \underline{f}_{\boldsymbol{X}}(\boldsymbol{x})\, e^{\boldsymbol{t}'\zeta^{view}(\boldsymbol{x})} d\boldsymbol{x} = \underline{\mathbb{E}}\{e^{\boldsymbol{t}'\zeta^{view}(\boldsymbol{X})}\}, \tag{A.3.2}$$

where $\underline{\mathbb{E}}\{\cdot\}$ denotes the expectation under the base distribution $\underline{f}_{\boldsymbol{X}}$ (1), or $\underline{\mathbb{E}}\{\cdot\} \equiv \mathbb{E}^{\underline{f}_{\boldsymbol{X}}}\{\cdot\}$.

A.3.2 Sample dual Lagrangian: gradient and Hessian

Due to the relationship $\hat{\psi}(\boldsymbol{t}; \{\underline{\mathcal{X}}, \underline{\boldsymbol{p}}\}) = \psi[\hat{\underline{f}}_{\boldsymbol{X}}](\boldsymbol{t})$ (3.22), the gradient and Hessian of the sample dual Lagrangian $\hat{\mathcal{L}}(\boldsymbol{t}; \boldsymbol{\eta}^{view}) \equiv \hat{\psi}(\boldsymbol{t}; \{\underline{\mathcal{X}}, \underline{\boldsymbol{p}}\}) - \boldsymbol{t}'\boldsymbol{\eta}^{view}$ can be derived as the sample of counterpart of the true dual Lagrangian $\mathcal{L}(\boldsymbol{t}; \boldsymbol{\eta}^{view}) \equiv \psi(\boldsymbol{t}) - \boldsymbol{t}'\boldsymbol{\eta}^{view}$ (28).

More precisely, the gradient is obtained by replacing the expectation in (29) with its sample counterpart, or

$$\begin{aligned} \nabla_{\boldsymbol{t}}\hat{\mathcal{L}}(\boldsymbol{t}; \boldsymbol{\eta}^{view}) &= \nabla_{\boldsymbol{t}}\hat{\psi}(\boldsymbol{t}; \{\underline{\mathcal{X}}, \underline{\boldsymbol{p}}\}) - \boldsymbol{\eta}^{view} \\ &= \widehat{\mathbb{E}}^{f_{\boldsymbol{t}}}\{\zeta^{view}(\boldsymbol{X})\} - \boldsymbol{\eta}^{view} \\ &= \textstyle\sum_{j=1}^{\bar{j}} p_{\boldsymbol{t}}^{(j)} \zeta^{view}(\underline{\boldsymbol{x}}^{(j)}) - \boldsymbol{\eta}^{view}, \end{aligned} \tag{A.3.3}$$

where $\boldsymbol{p}_{\boldsymbol{t}} = \mathit{softmax}(\ln(\underline{\boldsymbol{p}}) + \boldsymbol{t}'\underline{\mathcal{Z}})$ are defined as in (3.32).

Similarly, the Hessian is obtained by replacing the covariance in (30) with its sample counterpart, or

$$\begin{aligned}\nabla^2_{\boldsymbol{t},\boldsymbol{t}}\hat{\mathcal{L}}(\boldsymbol{t};\boldsymbol{\eta}^{view}) &= \nabla^2_{\boldsymbol{t},\boldsymbol{t}}\hat{\psi}(\boldsymbol{t};\{\underline{\mathcal{X}},\underline{\boldsymbol{p}}\}) \\ &= \widehat{\mathbb{C}v}^{f_{\boldsymbol{t}}}\{\zeta^{view}(\boldsymbol{X})\} \\ &= \textstyle\sum_{j=1}^{\bar{j}} p^{(j)}_{\boldsymbol{t}}(\zeta^{view}(\underline{\boldsymbol{x}}^{(j)})-\nabla_{\boldsymbol{t}}\hat{\psi}(\boldsymbol{t};\{\underline{\mathcal{X}},\underline{\boldsymbol{p}}\}))(\zeta^{view}(\underline{\boldsymbol{x}}^{(j)})-\nabla_{\boldsymbol{t}}\hat{\psi}(\boldsymbol{t};\{\underline{\mathcal{X}},\underline{\boldsymbol{p}}\}))',\end{aligned} \quad \text{(A.3.4)}$$

Hence, as long as $\zeta^{view}(\underline{\mathcal{X}}) = \{\zeta^{view}(\underline{\boldsymbol{x}}^{(j)})\}_{j=1}^{\bar{j}}$ (3.30) is a full rank matrix, the Hessian $\nabla^2_{\boldsymbol{t},\boldsymbol{t}}\hat{\mathcal{L}}(\boldsymbol{t};\boldsymbol{\eta}^{view})$ is a positive definite matrix and hence the sample dual Lagrangian $\hat{\mathcal{L}}(\boldsymbol{t};\boldsymbol{\eta}^{view})$ is a convex function.

A.3.3 Exponential family distributions: gradient of log-pdf

The generic n-th partial derivative of the log-pdf $u(\boldsymbol{x}) = \boldsymbol{t}'\zeta^{view}(\boldsymbol{x}) + \ln \underline{f}_{\boldsymbol{X}}(\boldsymbol{x})$ (3.17) reads

$$\begin{aligned}[\nabla_{\boldsymbol{x}}u(\boldsymbol{x})]_n &\equiv \frac{\partial}{\partial x_n}u_{\boldsymbol{t}}(\boldsymbol{x}) = \frac{\partial}{\partial x_n}[\textstyle\sum_{k=1}^{\bar{k}} t_k\zeta^{view}_k(\boldsymbol{x}) + \frac{\partial}{\partial x_n}\ln \underline{f}_{\boldsymbol{X}}(\boldsymbol{x})] \\ &= \textstyle\sum_{k=1}^{\bar{k}} t_k \frac{\partial}{\partial x_n}\zeta^{view}_k(\boldsymbol{x}) + \frac{1}{\underline{f}_{\boldsymbol{X}}(\boldsymbol{x})}\frac{\partial}{\partial x_n}\underline{f}_{\boldsymbol{X}}(\boldsymbol{x}) \\ &= \textstyle\sum_{k=1}^{\bar{k}} t_k[J_{\zeta^{view}}(\boldsymbol{x})]_{k,n} + \frac{1}{\underline{f}_{\boldsymbol{X}}(\boldsymbol{x})}[\nabla_{\boldsymbol{x}}\underline{f}_{\boldsymbol{X}}(\boldsymbol{x})]_n \\ &= [J_{\zeta^{view}}(\boldsymbol{x})'\boldsymbol{t}]_n + \frac{1}{\underline{f}_{\boldsymbol{X}}(\boldsymbol{x})}[\nabla_{\boldsymbol{x}}\underline{f}_{\boldsymbol{X}}(\boldsymbol{x})]_n,\end{aligned} \quad \text{(A.3.5)}$$

where in the third row we used the definition of Jacobian matrix

$$[J_{\zeta^{view}}(\boldsymbol{x})]_{k,n} \equiv \frac{\partial}{\partial x_n}\zeta^{view}_k(\boldsymbol{x}). \quad \text{(A.3.6)}$$

Hence comparing both sides of the above identity we obtain the desired result (3.18).

A.3.4 Dual Lagrangian: relationship with relative entropy

Since the updated distribution is an exponential family distribution $\boldsymbol{X} \sim \mathit{Exp}(\boldsymbol{\theta}^{view}, \zeta^{view}, \underline{f}_{\boldsymbol{X}})$ (3.3), it is immediate to verify that the relative entropy (17) between the base and the updated distribution is the negative dual Lagrangian (28)

$$\begin{aligned}\mathcal{E}(\bar{f}_{\boldsymbol{X}}\|\underline{f}_{\boldsymbol{X}}) &= \int_{\mathbb{R}^{\bar{n}}} f_{\boldsymbol{\theta}^{view}}(\boldsymbol{x})\ln(\frac{f_{\boldsymbol{\theta}^{view}}(\boldsymbol{x})}{\underline{f}_{\boldsymbol{X}}(\boldsymbol{x})})d\boldsymbol{x} \\ &= \int_{\mathbb{R}^{\bar{n}}} \underline{f}_{\boldsymbol{X}}(\boldsymbol{x})\, e^{\boldsymbol{\theta}^{view\prime}\zeta^{view}(\boldsymbol{x})-\psi(\boldsymbol{\theta}^{view})}(\boldsymbol{\theta}^{view\prime}\zeta^{view}(\boldsymbol{x})-\psi(\boldsymbol{\theta}^{view}))d\boldsymbol{x} \\ &= \boldsymbol{\theta}^{view\prime}\int_{\mathbb{R}^{\bar{n}}}\zeta^{view}(\boldsymbol{x})f_{\boldsymbol{\theta}^{view}}(\boldsymbol{x})\,d\boldsymbol{x}-\psi(\boldsymbol{\theta}^{view}) \\ &= \boldsymbol{\theta}^{view\prime}\boldsymbol{\eta}^{view}-\psi(\boldsymbol{\theta}^{view}) = -\mathcal{L}(\boldsymbol{\theta}^{view};\boldsymbol{\eta}^{view}),\end{aligned} \tag{A.3.7}$$

where in the second row we used the explicit expression of the updated pdf $\bar{f}_{\boldsymbol{X}}(\boldsymbol{x}) \equiv f_{\boldsymbol{\theta}^{view}}(\boldsymbol{x}) = \underline{f}_{\boldsymbol{X}}(\boldsymbol{x})\, e^{\boldsymbol{\theta}^{view\prime}\zeta^{view}(\boldsymbol{x})-\psi(\boldsymbol{\theta}^{view})}$ (24); and where in the last row we used the fact that the updated distribution by definition (16) satisfies the views (3.2), or $\mathbb{E}^{\bar{f}_{\boldsymbol{X}}}\left\{\zeta^{view}(\boldsymbol{X})\right\} = \boldsymbol{\eta}^{view}$.

A.3.5 Dual Lagrangian: gradient and Hessian with respect to features

First of all, let us consider the optimal Lagrange multipliers $\boldsymbol{\theta}^{view}$ (3.5) which are a suitable function θ^{view} of the features

$$\boldsymbol{\theta}^{view} = \theta^{view}(\boldsymbol{\eta}^{view}) \equiv \nabla_{\boldsymbol{t}}\psi^{-1}(\boldsymbol{\eta}^{view}), \tag{A.3.8}$$

also known as link function. See also [Amari and Nagaoka, 2000] and [Amari, 2016] for details.

From the chain rule, the gradient with respect to $\boldsymbol{\eta}^{view}$ of the minimal dual Lagrangian reads

$$\begin{aligned}\nabla_{\boldsymbol{\eta}^{view}}\mathcal{L}(\theta^{view}(\boldsymbol{\eta}^{view});\boldsymbol{\eta}^{view}) &= \nabla_{\boldsymbol{\eta}^{view}}[\psi(\theta^{view}(\boldsymbol{\eta}^{view}))-\theta^{view}(\boldsymbol{\eta}^{view})'\boldsymbol{\eta}^{view}] \\ &= (J_{\theta^{view}}(\boldsymbol{\eta}^{view}))'[\nabla_{\boldsymbol{t}}\psi(\theta(\boldsymbol{\eta}^{view}))-\boldsymbol{\eta}^{view}]-\theta^{view}(\boldsymbol{\eta}^{view}) \\ &= -\theta^{view}(\boldsymbol{\eta}^{view}),\end{aligned} \tag{A.3.9}$$

as follows from the fact that by definition of Lagrange multipliers $\theta^{view}(\boldsymbol{\eta}^{view})$ (27)

$$\nabla_{\boldsymbol{t}}\psi(\theta^{view}(\boldsymbol{\eta}^{view}))-\boldsymbol{\eta}^{view} = \boldsymbol{0}_{\bar{k}\times 1}. \tag{A.3.10}$$

This easily implies the expression of the views intensity (3.52) from the relationship between minimal relative entropy and dual Lagrangian [A.3.4].

Moreover, from the chain rule the Hessian with respect to $\boldsymbol{\eta}^{view}$ of the minimal dual Lagrangian reads

$$\begin{aligned}\nabla^2_{\boldsymbol{\eta}^{view},\boldsymbol{\eta}^{view}}\mathcal{L}(\theta^{view}(\boldsymbol{\eta}^{view});\boldsymbol{\eta}^{view}) &= (J_{\theta^{view}}(\boldsymbol{\eta}^{view}))'\times\nabla^2_{\boldsymbol{t},\boldsymbol{t}}\mathcal{L}(\theta^{view}(\boldsymbol{\eta}^{view});\boldsymbol{\eta}^{view})\times(J_{\theta^{view}}(\boldsymbol{\eta}^{view}))' \\ &= (\nabla^2_{\boldsymbol{t},\boldsymbol{t}}\psi(\theta^{view}(\boldsymbol{\eta}^{view})))^{-1}\end{aligned} \tag{A.3.11}$$

as follows from the fact that the $\bar{k}\times\bar{k}$ Jacobian matrix of $\theta^{view}(\boldsymbol{\eta}^{view})$ reads

$$J_{\theta^{view}}(\boldsymbol{\eta}^{view}) = (\nabla^2_{\boldsymbol{t},\boldsymbol{t}}\psi(\theta^{view}(\boldsymbol{\eta}^{view})))^{-1}, \tag{A.3.12}$$

because of the inverse function differentiation; and $\nabla^2_{\boldsymbol{t},\boldsymbol{t}}\mathcal{L}(\theta^{view}(\boldsymbol{\eta}^{view});\boldsymbol{\eta}^{view}) = \nabla^2_{\boldsymbol{t},\boldsymbol{t}}\psi(\theta^{view}(\boldsymbol{\eta}^{view}))$ (30).

A.3.6 ENS: relationship with relative entropy

Let us consider the discrete differential Shannon entropy, which for a generic vector of probabilities reads

$$\mathcal{H}(\boldsymbol{p}) \equiv -\sum_{j=1}^{\bar{j}} p^{(j)}\ln p^{(j)}. \tag{A.3.13}$$

Then, let us consider the exponential probabilities $\bar{\boldsymbol{p}} \equiv \mathit{softmax}(\ln(\underline{\boldsymbol{p}}) + \hat{\boldsymbol{\theta}}^{view\prime}\underline{\mathcal{Z}})$ (3.32), then the generic case (3.55) will follow similarly.

We have the following identities hold

$$\begin{aligned}
\mathcal{H}(\bar{\boldsymbol{p}}) &\equiv -\sum_{j=1}^{\bar{j}} \bar{p}^{(j)}\ln \bar{p}^{(j)} \\
&= -\sum_{j=1}^{\bar{j}} \bar{p}^{(j)}[\ln \underline{p}^{(j)} + \hat{\boldsymbol{\theta}}^{view\prime}\zeta^{view}(\underline{\boldsymbol{x}}^{(j)}) - \hat{\psi}(\hat{\boldsymbol{\theta}}^{view};\{\underline{\mathcal{X}},\underline{\boldsymbol{p}}\})] \\
&= \ln\bar{j}(\sum_{j=1}^{\bar{j}}\bar{p}^{(j)}) - \hat{\boldsymbol{\theta}}^{view\prime}\sum_{j=1}^{\bar{j}}\bar{p}^{(j)}\zeta^{view}(\underline{\boldsymbol{x}}^{(j)}) + \hat{\psi}(\hat{\boldsymbol{\theta}}^{view};\{\underline{\mathcal{X}},\underline{\boldsymbol{p}}\}) \\
&= \ln\bar{j} - \hat{\boldsymbol{\theta}}^{view\prime}\boldsymbol{\eta}^{view} + \hat{\psi}(\hat{\boldsymbol{\theta}}^{view};\{\underline{\mathcal{X}},\underline{\boldsymbol{p}}\}) \\
&= \ln\bar{j} - \mathcal{E}(\bar{\boldsymbol{p}}||\underline{\boldsymbol{p}}),
\end{aligned} \tag{A.3.14}$$

where in the fourth row we used the fact that $\sum_{j=1}^{\bar{j}}\bar{p}^{(j)}\zeta^{view}(\underline{\boldsymbol{x}}^{(j)}) = \boldsymbol{\eta}^{view}$ (3.34); and in the last row we used $\mathcal{E}(\bar{\boldsymbol{p}}||\underline{\boldsymbol{p}}) \equiv -\hat{\mathcal{L}}(\hat{\boldsymbol{\theta}}^{view};\boldsymbol{\eta}^{view}) = -\hat{\psi}(\hat{\boldsymbol{\theta}}^{view};\{\underline{\mathcal{X}},\underline{\boldsymbol{p}}\}) + \hat{\boldsymbol{\theta}}^{view\prime}\boldsymbol{\eta}^{view}$ (which follows using similar arguments as for the true counterpart [A.3.4])

Then by taking the exponential of (A.3.13) we obtain

$$e^{\mathcal{H}(\bar{\boldsymbol{p}})} = \bar{j}e^{-\mathcal{E}(\bar{\boldsymbol{p}}||\underline{\boldsymbol{p}})}, \tag{A.3.15}$$

Finally, because of the law of large numbers we can approximate

$$\mathcal{E}(\bar{\boldsymbol{p}}||\underline{\boldsymbol{p}}) \approx \mathcal{E}(\bar{f}_{\boldsymbol{X}}\|\underline{f}_{\boldsymbol{X}}), \tag{A.3.16}$$

which implies the desired result (3.58).

A.3.7 Exponential family distributions: iterative update

Suppose that $f_{\boldsymbol{X}}$ belongs to the exponential family class (23) under the base distribution $\underline{f}_{\boldsymbol{X}}$(1),

$$f_{\boldsymbol{X}} \sim \mathit{Exp}(\boldsymbol{t},\zeta^{view},\underline{f}_{\boldsymbol{X}}), \tag{A.3.17}$$

for some arbitrary $\bar{k} \times 1$ vector of Lagrange multipliers $\boldsymbol{t} \equiv (t_1, \ldots, t_{\bar{k}})'$.

Then, according to the conjugate property (A.1.9), given any other vector $\boldsymbol{\theta}^{view} \equiv (\theta_1^{view}, \ldots, \theta_{\bar{k}}^{view})'$, the exponential family distribution $\bar{f}_{\boldsymbol{X}} \sim Exp(\boldsymbol{\theta}^{view}, \zeta^{view}, f_{\boldsymbol{X}})$ (23) is also an exponential family distribution (23) under $\underline{f}_{\boldsymbol{X}}$ as reference measure, or

$$\bar{f}_{\boldsymbol{X}} \sim Exp(\boldsymbol{t} + \boldsymbol{\theta}^{view}, \zeta^{view}, \underline{f}_{\boldsymbol{X}}). \tag{A.3.18}$$

This also means from (A.1.10)

$$\psi_{\underline{f}_{\boldsymbol{X}}, \zeta^{view}}(\boldsymbol{t} + \boldsymbol{\theta}^{view}) = \psi_{\underline{f}_{\boldsymbol{X}}, \zeta^{view}}(\boldsymbol{t}) + \psi_{f_{\boldsymbol{X}}, \zeta^{view}}(\boldsymbol{\theta}^{view}). \tag{A.3.19}$$

Then by induction, given any arbitrary sequence of Lagrange multipliers $\boldsymbol{\theta}^{view(i)}$, where $\boldsymbol{\theta}^{view(0)} \equiv \mathbf{0}$, and defining a recursive sequence of exponential family distributions (23)

$$f_{\boldsymbol{X}}^{(i)} \sim Exp(\boldsymbol{\theta}^{view(i)}, \zeta^{view}, f_{\boldsymbol{X}}^{(i-1)}), \tag{A.3.20}$$

where $\underline{f}_{\boldsymbol{X}}^{(0)} \equiv \underline{f}_{\boldsymbol{X}}$, it is immediate to verify that the following holds

$$f_{\boldsymbol{X}}^{(\bar{\imath})} \sim Exp(\boldsymbol{\theta}^{view}, \zeta^{view}, \underline{f}_{\boldsymbol{X}}) \tag{A.3.21}$$

where

$$\boldsymbol{\theta}^{view} \equiv \textstyle\sum_{i=1}^{\bar{\imath}} \boldsymbol{\theta}^{view(i)}. \tag{A.3.22}$$

A.3.8 Exponential family distributions: invariance of the updated distribution

Suppose that our base distribution (1) is an exponential family distribution (23)

$$\underline{f}_{\boldsymbol{X}} \sim Exp(\underline{\boldsymbol{\theta}}, \zeta^{view}, h), \tag{A.3.23}$$

for some base vector $\underline{\boldsymbol{\theta}} \equiv (\underline{\theta}_1, \ldots, \underline{\theta}_{\bar{k}})'$ of canonical coordinates and arbitrary reference measure $h(\boldsymbol{x}) > 0$, which without loss of generality we can assume to be normalized $\int h(\boldsymbol{x}) d\boldsymbol{x} = 1$.

Then, under views on generalized expectations $\mathcal{C}_{\boldsymbol{X}} : \mathbb{E}^{f_{\boldsymbol{X}}} \left\{\zeta^{view}(\boldsymbol{X})\right\} = \boldsymbol{\eta}^{view}$ as in (3.2), the updated $\bar{f}_{\boldsymbol{X}}$ (16) is an exponential family distribution (3.3) of the following form

$$\bar{f}_{\boldsymbol{X}} \sim Exp(\boldsymbol{\theta}^{view}, \zeta^{view}, \underline{f}_{\boldsymbol{X}}), \tag{A.3.24}$$

where the optimal Lagrange multipliers $\boldsymbol{\theta}^{view} \equiv (\theta_1^{view}, \ldots, \theta_{\bar{k}}^{view})'$ are the solutions of the dual Lagrangian problem (3.5)

$$\boldsymbol{\theta}^{view} \equiv \operatorname*{argmin}_{\boldsymbol{t}} \psi_{\underline{f}_{\boldsymbol{X}}, \zeta^{view}}(\boldsymbol{t}) - \boldsymbol{t}'\boldsymbol{\eta}^{view}, \tag{A.3.25}$$

where $\psi_{\underline{f}_{\boldsymbol{X}}, \zeta^{view}}(\boldsymbol{t}) \equiv \ln \int_{\mathbb{R}^{\bar{n}}} e^{\boldsymbol{t}'\zeta^{view}(\boldsymbol{x})} \underline{f}_{\boldsymbol{X}}(\boldsymbol{x})\, d\boldsymbol{x}$ is the log-partition function (25)

Moreover, from the property of composition of exponential family distributions [A.3.7], the updated $\bar{f}_{\boldsymbol{X}}$ (16) can be expressed also as an exponential family distribution under the reference measure h, or

$$\bar{f}_{\boldsymbol{X}} \sim Exp(\underline{\boldsymbol{\theta}} + \boldsymbol{\theta}^{view}, \zeta^{view}, h). \tag{A.3.26}$$

Finally, it is easy to verify that $\bar{f}_{\boldsymbol{X}}$ (A.3.26) is also the updated distribution (3.3) under the *same* views $\mathcal{C}_{\boldsymbol{X}}$ on generalized expectations as above and reference measure h as base distribution (1).

To this purpose, we just need to verify that the vector $\underline{\boldsymbol{\theta}} + \boldsymbol{\theta}^{view}$ is the solution of the dual Lagrangian problem (3.5)

$$\underline{\boldsymbol{\theta}} + \boldsymbol{\theta}^{view} = \bar{\boldsymbol{\vartheta}} \equiv \underset{\boldsymbol{\vartheta}}{\operatorname{argmin}}\, \psi^{h}(\boldsymbol{\vartheta}) - \boldsymbol{\vartheta}'\boldsymbol{\eta}^{view}, \tag{A.3.27}$$

where $\psi_{h,\zeta^{view}}(\boldsymbol{\vartheta}) \equiv \ln \int_{\mathbb{R}^{\bar{n}}} e^{\boldsymbol{\vartheta}'\zeta^{view}(\boldsymbol{x})} h\,(\boldsymbol{x})\, d\boldsymbol{x}$ denotes the log-partition function as in (25) under the reference measure h.

Indeed, the following two optimizations are equivalent

$$\begin{aligned} \boldsymbol{\theta}^{view} &\equiv \underset{\boldsymbol{t}}{\operatorname{argmin}}\, \psi_{\underline{f}_{\boldsymbol{X}},\zeta^{view}}(\boldsymbol{t}) - \boldsymbol{t}'\boldsymbol{\eta}^{view} \\ &= \underset{\boldsymbol{t}}{\operatorname{argmin}}\, \psi_{h,\zeta^{view}}(\underline{\boldsymbol{\theta}} + \boldsymbol{t}) - \psi_{h,\zeta^{view}}(\underline{\boldsymbol{\theta}}) - \boldsymbol{t}'\boldsymbol{\eta}^{view} \\ &= \underset{\boldsymbol{t}}{\operatorname{argmin}}\, \psi_{h,\zeta^{view}}(\underline{\boldsymbol{\theta}} + \boldsymbol{t}) - (\underline{\boldsymbol{\theta}} + \boldsymbol{t})'\boldsymbol{\eta}^{view}, \end{aligned} \tag{A.3.28}$$

where the second row follows from (A.3.19)

$$\psi_{\underline{f}_{\boldsymbol{X}},\zeta^{view}}(\boldsymbol{t}) = \psi_{h,\zeta^{view}}(\underline{\boldsymbol{\theta}} + \boldsymbol{t}) - \psi_{h,\zeta^{view}}(\underline{\boldsymbol{\theta}}); \tag{A.3.29}$$

and the last row follows because the constant terms $\psi_{h,\zeta^{view}}(\underline{\boldsymbol{\theta}})$ and $\underline{\boldsymbol{\theta}}'\boldsymbol{\eta}^{view}$ do not alter the optimization problem.

Hence, changing the coordinates in (A.3.27) via shifting

$$\boldsymbol{\vartheta} \equiv \underline{\boldsymbol{\theta}} + \boldsymbol{t}, \tag{A.3.30}$$

we obtain the desired result (A.3.26).

A.3.9 Convergence of the iterative approach

Because of the invariance of the updated distribution (3.43), for a given i-th base distribution $f_{\boldsymbol{X}}^{(i)}$ (3.44), the true updated distribution $\bar{f}_{\boldsymbol{X}}$ (3.3) reads

$$\bar{f}_{\boldsymbol{X}}(\boldsymbol{x}) = f_{\boldsymbol{X}}^{(i)}(\boldsymbol{x}) e^{\boldsymbol{\epsilon}^{view(i+1)\prime}\zeta^{view}(\boldsymbol{x}) - \psi^{(i)}(\boldsymbol{\epsilon}^{view(i+1)})}, \tag{A.3.31}$$

for some vector of optimal Lagrange multipliers $\boldsymbol{\epsilon}^{view(i+1)}$ (3.5) and where $\psi^{(i)}$ is the i-th log-partition function (25)

$$\psi^{(i)}(\boldsymbol{\epsilon}^{view(i+1)}) \equiv \psi_{f_{\boldsymbol{X}}^{(i)},\zeta^{view}}(\boldsymbol{\epsilon}^{view(i+1)}).$$

Now, let us assume that the infinite sum of Lagrange multipliers is finite

$$\sum_{i=1}^{\infty} \boldsymbol{\epsilon}^{view(i)} < \infty, \tag{A.3.32}$$

and hence that the sequence $\boldsymbol{\epsilon}^{view(i)}$ converges to zero sufficiently fast

$$||\boldsymbol{\epsilon}^{view(i)}|| \underset{i\to\infty}{\to} 0. \tag{A.3.33}$$

Then, it turns out that the sequence $f_{\boldsymbol{X}}^{(i)}$ converges to $\bar{f}_{\boldsymbol{X}}$ (3.46).

Indeed, according to [A.3.4], the relative entropy (17) between the true updated distribution $\bar{f}_{\boldsymbol{X}}$ and i-th base distribution $f_{\boldsymbol{X}}^{(i)}$ reads

$$\mathcal{E}(\bar{f}_{\boldsymbol{X}} \| f_{\boldsymbol{X}}^{(i)}) = \boldsymbol{\epsilon}^{view(i+1)\prime}\boldsymbol{\eta}^{view} - \psi^{(i)}(\boldsymbol{\epsilon}^{view(i+1)}), \tag{A.3.34}$$

and hence must converge to zero

$$\mathcal{E}(\bar{f}_{\boldsymbol{X}} \| f_{\boldsymbol{X}}^{(i)}) \underset{i\to\infty}{\to} 0, \tag{A.3.35}$$

for continuity, since we have for any i

$$\psi^{(i)}(\mathbf{0}) = 0. \tag{A.3.36}$$