Remarks on the Interpolation Method

Roberto Boccagna[1] · **Davide Gabrielli**[1]

Abstract

We discuss a generalization of the classic condition of validity of the interpolation method for the density of quenched free energy of mean field spin glasses. The condition is written just in terms of the L^2 metric structure of the Gaussian random variables. As an example of application we deduce the existence of the thermodynamic limit for a GREM model with infinite branches for which the classic conditions of validity fail. We underline the dependence of the density of quenched free energy just on the metric structure and discuss the models from a metric viewpoint.

Keywords Spin glasses · Interpolation method · Thermodynamic limit

Mathematics Subject Classification 60G15 · 82D30

1 Introduction

The interpolation method is a simple but powerful technique used to prove inequalities for Gaussian random vectors (see for example [20] and [21]). This method has great relevance in the field of Mathematical and Theoretical Physics since it represents an essential ingredient in the study of mean field spin glasses. In the breakthrough paper [19] it has been used to prove the existence of the thermodynamic limit for the quenched density of free energy for the Sherrington–Kirkpatrick model. This was a longstanding problem and its solution was the turning point towards the proof of the Parisi Formula [29].

Spin glasses are simple mathematical models for disordered systems whose rigorous analysis is indeed a *challenge for mathematicians*. We refer to [24,28] the mathematically interested reader and to [23] the physically interested one. Among plenty of models, one of the most studied is that introduced by Sherrington and Kirkpatrick in [26] as a solvable elementary model. Indeed the structure of the solution turned out to be much more rich and complex than expected and was build up in a series of papers by Parisi (see [23] for a detailed

Communicated by Federico Ricci-Tersenghi.

✉ Davide Gabrielli
davide.gabrielli@univaq.it

1 Università dell'Aquila, Via Vetoio, Loc. Coppito, 67010 L'Aquila, Italy

discussion). A rigorous proof of the Parisi conjectured solution was missing for a long time and the interpolation method played a key role in its proof. See [18] for a review on this.

Using the same idea of [19], the authors of [12] proposed a general setting for the interpolation method in the framework of mean field spin glasses. Furthermore, they successfully applied this technique to prove the existence of the thermodynamic limit for the *Generalized Random Energy Model* (GREM, a family of models introduced in [15]) with a finite number of levels.

The interpolation method is now a powerful technique that has many different applications in different contexts, see for example [1–6,22], a list that is by far not exaustive.

The "classical" hypothesis under which the interpolation method can be applied to the quenched free energy of mean field spin glasses consists of a collection of equalities and inequalities for the covariance matrix of the underlying multivariate Gaussian process. We show that less restrictive conditions are actually needed. More precisely, we show that the method works under conditions that involve just the L^2 metric structure of the Gaussian random vectors. By the correspondence in [27], [17] this is always an Euclidean metric structure. A condition of this type is very natural since the quenched free energy depends on the distribution of the Gaussian random vector only through its metric structure. This gives an interesting geometric flavor and interpretation and we discuss, at the end of the paper, the models from a purely metric viewpoint. This generalized condition of validity was also obtained through a tricky computation in the framework of Sudakov-Fernique inequalities in [11]. Here we deduce the condition by a general argument that could in principle be applied also to comparative inequalities involving expected values of different functions of Gaussian vectors. As an example of application of the generalized condition, we consider a GREM model with infinite levels and deduce the existence of the thermodynamic limit for the quenched density of free energy. Indeed, in this case the usual condition of validity of the interpolation method used in [12,19] fails. We can deduce therefore the existence of the thermodynamic limit directly using the simple argument of the interpolation method. We refer to [9,25] and [10] for the beautiful mathematics involved in the limit of such kind of models.

The structure of the paper is the following.

In Sect. 2 we briefly recall the basics of the interpolation method together with the conditions used in [19] and [12]; we then discuss the Euclidean metric structure associated to any Gaussian random vector and finally show the generalized conditions.

In Sect. 3 we discuss two examples. The first one is the Sherrington–Kirkpatrick model. This is done simply to recall the basic mechanism and idea of application. The second example is a GREM model with infinite levels for which it is necessary to use the generalized conditions to prove the existence of the thermodynamic limit. In the final part of this section we discuss the models from a purely metric viewpoint introducing a class of models that have a natural metric structure and for which it is possible to show the existence of the thermodynamic limit.

In the Appendix we have an elementary auxiliary Lemma.

2 The Interpolation Method

2.1 The Interpolation Method

Let $X = (X_1, \ldots, X_n)$ be a n-dimensional zero mean Gaussian random vector having covariance matrix C. The $n \times n$ symmetric matrix C is non-negative definite and the elements are defined by $C_{i,j} := \mathbb{E}[X_i X_j]$. When C is positive definite then the distribution of X is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^n$ and the density is

$$\phi_C(x) := \frac{1}{\sqrt{(2\pi)^n \det(C)}} e^{-\frac{1}{2}(x, C^{-1}x)}, \tag{2.1}$$

where $(\cdot, \cdot)$ denotes the Euclidean scalar product in $\mathbb{R}^n$. We restrict to the case of positive definite matrices since the other cases can be deduced by a limiting procedure. We have the Fourier transform representation

$$\phi_C(x) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} d\lambda \, e^{-i(\lambda, x)} e^{-\frac{1}{2}(\lambda, C\lambda)}. \tag{2.2}$$

We denote by $\mathrm{Tr}(\cdot)$ the trace of a matrix and consider $\bar{\mathcal{C}}$ the set of non-negative definite symmetric matrices endowed with the Hilbert-Schmidt scalar product

$$(A, B) := \mathrm{Tr}(AB), \qquad A, B \in \bar{\mathcal{C}}. \tag{2.3}$$

The open set of positive definite symmetric matrices corresponds to $\mathcal{C}$.

Let $\phi : \mathcal{C} \times \mathbb{R}^n \to \mathbb{R}^+$ as defined in (2.1). By (2.2) and a direct computation we have

$$\frac{\partial \phi_C(x)}{\partial C_{i,j}} = \frac{\partial \phi_C(x)}{\partial C_{j,i}} = \frac{\partial^2 \phi_C(x)}{\partial x_i \partial x_j}, \tag{2.4}$$

and

$$\frac{\partial \phi_C(x)}{\partial C_{i,i}} = \frac{1}{2} \frac{\partial^2 \phi_C(x)}{\partial x_i^2}. \tag{2.5}$$

Recall that in the above formulas C is a symmetric matrix so that the variations in the computation of (2.4) are constructed varying symmetrically the matrix C. More precisely let $E^{\{i,j\}}$ with $i \neq j$ be the symmetric matrix such that $E_{i,j}^{\{i,j\}} = E_{j,i}^{\{i,j\}} = 1$ and having all the remaining elements equal to zero. Given $F : \mathcal{C} \to \mathbb{R}$ we define

$$\frac{\partial F(C)}{\partial C_{j,i}} = \frac{\partial F(C)}{\partial C_{i,j}} := \lim_{\delta \to 0} \frac{F(C + \delta E^{\{i,j\}}) - F(C)}{\delta}. \tag{2.6}$$

Consider now $f : \mathbb{R}^n \to \mathbb{R}$ a C^2 function with *moderate growth at infinity*, for example such that $|f(x)| \leq e^{\lambda |x|}$ for a suitable constant $\lambda \geq 0$. This technical condition is related to the validity of some integrations by parts. We call $\nabla^2 f(x)$ the Hessian matrix of f at x, that is the symmetric matrix having elements

$$\left(\nabla^2 f\right)_{i,j}(x) := \frac{\partial^2 f(x)}{\partial x_i \partial x_j}.$$

The following result is the interpolation method. For the readers convenience we give the short proof.

Lemma 2.1 (Interpolation method) *Consider two zero mean Gaussian random vectors X, Y having covariance matrices respectively given by C^X and C^Y. Consider a C^2 function f with moderate growth. We have*

$$\mathbb{E}\left[f\left(Y\right)\right] - \mathbb{E}\left[f\left(X\right)\right] = \frac{1}{2}\int_0^1 \mathrm{d}t\,\mathbb{E}\left[\mathrm{Tr}\left(\left(C^Y - C^X\right)\nabla^2 f\left(Z\left(t\right)\right)\right)\right], \quad (2.7)$$

where

$$Z(t) = \sqrt{t}X + \sqrt{(1-t)}Y, \quad (2.8)$$

and X, Y are two independent copies of the random vectors.

Proof When Z is a n-dimensional centered Gaussian random vector, then $\mathbb{E}\left[f(Z)\right]$ depends only on the covariance matrix C of the vector Z. Fix a C^2 function f and define the function $F : \bar{\mathcal{C}} \to \mathbb{R}$ as

$$F(C) := \mathbb{E}\left[f\left(Z\right)\right]. \quad (2.9)$$

With the help of formulas (2.4), (2.5), when $C \in \mathcal{C}$ we can compute

$$\frac{\partial F\left(C\right)}{\partial C_{i,j}} = \int_{\mathbb{R}^n} \mathrm{d}x\,\frac{\partial \phi_C\left(x\right)}{\partial C_{i,j}} f\left(x\right) = \int_{\mathbb{R}^n} \mathrm{d}x\,\frac{\partial^2 \phi_C\left(x\right)}{\partial x_i \partial x_j} f\left(x\right) \quad (2.10)$$

$$= \int_{\mathbb{R}^n} \mathrm{d}x\,\phi_C\left(x\right)\frac{\partial^2 f\left(x\right)}{\partial x_i \partial x_j} = \mathbb{E}\left[\left(\nabla^2 f\right)_{i,j}\left(Z\right)\right], \quad (2.11)$$

and

$$\frac{\partial F\left(C\right)}{\partial C_{i,i}} = \int_{\mathbb{R}^n} \mathrm{d}x\,\frac{\partial \phi_C\left(x\right)}{\partial C_{i,i}} f\left(x\right) = \frac{1}{2}\int_{\mathbb{R}^n} \mathrm{d}x\,\frac{\partial^2 \phi_C\left(x\right)}{\partial x_i^2} f\left(x\right)$$

$$= \frac{1}{2}\int_{\mathbb{R}^n} \mathrm{d}x\,\phi_C\left(x\right)\frac{\partial^2 f\left(x\right)}{\partial x_i^2} = \frac{1}{2}\mathbb{E}\left[\left(\nabla^2 f\right)_{i,i}\left(Z\right)\right]. \quad (2.12)$$

Given a C^1 parametric curve $\{C(t)\}_{t\in[0,1]}$ on $\mathcal{C}$ such that $C(0) = C^X$ and $C(1) = C^Y$, then we have

$$\mathbb{E}\left[f\left(Y\right)\right] - \mathbb{E}\left[f\left(X\right)\right] = \frac{1}{2}\int_0^1 \mathrm{d}t\,\mathbb{E}\left[\mathrm{Tr}\left(\frac{\mathrm{d}C\left(t\right)}{\mathrm{d}t}\nabla^2 f\left(Z\left(t\right)\right)\right)\right], \quad (2.13)$$

where $Z\left(t\right)$ is a centered Gaussian random vector having covariance $C\left(t\right)$. The special case when the curve linearly interpolates between C^X and C^Y gives (2.7) with $Z(t)$ given by (2.8). If one or both the matrices C^X and C^Y are not strictly positive definite, it is possible to add to the matrices $\varepsilon\mathbb{I}$, do the same computation as above and finally take the limit $\varepsilon \to 0$. □

The above formula is the core of the interpolation method. It is very useful to establish inequalities between the two expected values on the left hand side of (2.7).

The *Guerra–Toninelli interpolation method* is a simple but powerful technique developed in the study of mean field spin glasses (see [18,19] and references therein), which is based on an abstract theorem about Gaussian random variables. It corresponds to the interpolation method Lemma 2.1 with the special choice of the function

$$f(x) = \log \sum_{i=1}^n w_i \mathrm{e}^{x_i}, \quad (2.14)$$

where $w_i \in \mathbb{R}^+$ are some fixed positive weights.

In particular, Guerra and Toninelli obtained and used the following result (this is Theorem 2 in [18]) to prove the existence of the thermodynamic limit of the Sherrington–Kirkpatrick model. The same idea and the same Theorem (Theorem 2.2 below), was used later on in [12] to deduce the existence of the thermodynamic limit for a GREM model [15] with a finite number of levels.

Theorem 2.2 *Let X, Y two centered Gaussian random vectors and the function f given by* (2.14). *If*

$$C_{i,i}^X = C_{i,i}^Y, \qquad \forall i, \tag{2.15}$$

$$C_{i,j}^X \geq C_{i,j}^Y, \qquad \forall i \neq j, \tag{2.16}$$

then we have

$$\mathbb{E}\left[f\left(Y\right)\right] \geq \mathbb{E}\left[f\left(X\right)\right]. \tag{2.17}$$

We show the proof of Theorem 2.2 that is based on the interpolation formula (2.7).

Proof of Theorem 2.2 Let us call, for any $i = 1, \ldots, n$

$$\mu_i\left(x\right) := \frac{w_i \mathrm{e}^{x_i}}{\sum_{j=1}^n w_j \mathrm{e}^{x_j}}. \tag{2.18}$$

By a direct computation, when f is (2.14), we have

$$\frac{\partial^2 f\left(x\right)}{\partial x_i^2} = \mu_i\left(x\right) - \mu_i^2\left(x\right), \tag{2.19}$$

$$\frac{\partial^2 f\left(x\right)}{\partial x_i \partial x_j} = -\mu_i\left(x\right)\mu_j\left(x\right). \tag{2.20}$$

By the formulas (2.19), (2.20) and conditions (2.15), (2.16), we have that

$$(C^Y - C^X)_{i,j}\left(\nabla^2 f\right)_{i,j}(x) \geq 0, \qquad \forall x \in \mathbb{R}^d, \quad \forall i, j \tag{2.21}$$

and the result follows by (2.7).   □

2.2 Covariances and Metrics

We start recalling some simple but useful Lemmas.

Lemma 2.3 *We have that the $n \times n$ symmetric matrix C belongs to $\mathcal{C}$ if and only if there exist n vectors $a^{(i)} \in \mathbb{R}^n$ such that*

$$C_{i,j} = \left(a^{(i)}, a^{(j)}\right). \tag{2.22}$$

This is a classic result and the matrix C is called the *Gram matrix* of the vectors $\left(a^{(i)}\right)_{i=1,\ldots,n}$, see for example [7].

A finite metric space with n points is called *Euclidean* if there exists a collection of n points on $\mathbb{R}^k$ having the same relative interdistances. Of course we can always fix $k = n$. Not every metric space can be realized in this way. The simplest example is the minimal path metrics on the vertices of the graph in Fig. 1 where the edges have all length 1.

Given a centered Gaussian random vector X there is naturally associated the metric d_X that is the L^2 distance between the random variables

$$d_X(i, j) := \sqrt{\mathbb{E}\left[\left(X_i - X_j\right)^2\right]} = \sqrt{C_{i,i}^X + C_{jj}^X - 2C_{i,j}^X}. \tag{2.23}$$
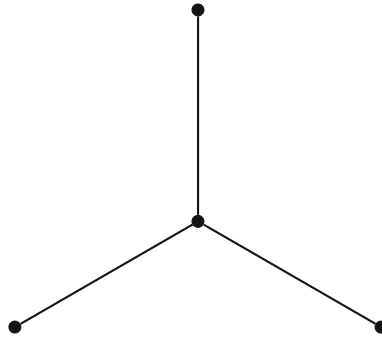
Fig. 1 An example of a non Euclidean metric spcace. The distance is the minimal path distance on the graph

We have the following result (see also [17,27])

Lemma 2.4 *A finite metric space* $(\{1, \ldots, n\}, d)$ *is Euclidean if and only if there exists a zero mean Gaussian random vector* $X = (X_1, \ldots, X_n)$ *such that* $d = d_X$.

Proof Consider d an Euclidean distance and let $a^{(i)}$, $i = 1, \ldots, n$ be some points on $\mathbb{R}^n$ that realize such a distance. This means that $d(i, j) = |a^{(i)} - a^{(j)}|$ where $|\cdot|$ represents the Euclidean distance in $\mathbb{R}^n$. Such a collection of vectors exists by definition of an Euclidean metric space. Let A be an $n \times n$ matrix defined by $A_{i,j} := a_j^{(i)}$. Let $Z = (Z_1, \ldots, Z_n)$ be a vector of i.i.d. standard Gaussian random variables and consider the Gaussian vector $X = AZ$ whose covariance $C^X = AA^T$ coincides with the right hand side of (2.22). Using (2.23) we have

$$d_X(i, j) = |a^{(i)} - a^{(j)}| = d(i, j). \tag{2.24}$$

Conversely let X a Gaussian zero mean vector with covariance C^X and let A an $n \times n$ matrix such that $C^X = AA^T$. Define n vectors in $\mathbb{R}^n$ by $a_j^{(i)} := A_{i,j}$; by (2.23) we have that d_X is determined by the first equality in (2.24) and is therefore Euclidean. □

Other simple but useful lemmas are the following. We give just the statements, the proofs can be found for example in [8].

Lemma 2.5 *Let* $v^{(1)}, \ldots, v^{(n)}$ *and* $w^{(1)}, \ldots, w^{(n)}$ *be two collections of n vectors in* $\mathbb{R}^n$. *We have that*

$$\left(v^{(i)}, v^{(j)}\right) = \left(w^{(i)}, w^{(j)}\right) \quad \forall i, j \tag{2.25}$$

if and only if there exists $O \in O(n)$ *such that* $w^{(i)} = O v^{(i)}$ *for any i.*

Lemma 2.6 *Let* $v^{(1)}, \ldots, v^{(n)}$ *and* $w^{(1)}, \ldots, w^{(n)}$ *be two collections of vectors in* $\mathbb{R}^n$. *We have that*

$$\left|v^{(i)} - v^{(j)}\right| = \left|w^{(i)} - w^{(j)}\right|, \quad \forall i, j, \tag{2.26}$$

if and only if there exists $O \in O(n)$ *and a vector* $b \in \mathbb{R}^n$ *such that*

$$w^{(i)} = O v^{(i)} + b, \quad \forall i. \tag{2.27}$$

The metric structure d_X, associated to a Gaussian random vector X, contains less information than the covariance C^X and there are random vectors having different covariances but the same metric structure. This type of invariance is best understood in terms of the vectors in $\mathbb{R}^n$ using the above Lemmas that characterize invariance by rotations and translations. In particular we can completely characterize the centered Gaussian random variables that share the same metric structure.

Lemma 2.7 *Given X and Y two n-dimensional centered Gaussian random vectors, we have that $d_X = d_Y$ if and only if there exists a centered Gaussian random variable W such that the random vector $X_i + W$, $i = 1, \ldots, n$ has the same distribution of Y.*

Proof If Y has the same distribution of $X + W 1_n$, where 1_n is the n-dimensional vector of all ones $1_n = (1, 1, \ldots, 1)$, then

$$d_Y(i, j) = \sqrt{\mathbb{E}\left[(Y_i - Y_j)^2 \right]} = \sqrt{\mathbb{E}\left[(X_i + W - X_j - W)^2 \right]} = d_X(i, j).$$

Conversely, suppose that $d_X = d_Y$. We have that there exist two matrices A^X and A^Y such that $A^X Z$ has the same distribution of X and $A^Y Z$ has the same distribution of Y, where Z is an n vector of i.i.d. standard Gaussian random variables. We define two collections $v^{(i)}$, $w^{(i)}$, $i = 1, \ldots n$ of vectors in $\mathbb{R}^n$ defined by $v_j^{(i)} := A_{i,j}^X$ and $w_j^{(i)} := A_{i,j}^Y$. Since $d_X = d_Y$ we have

$$\left| v^{(i)} - v^{(j)} \right| = \left| w^{(i)} - w^{(j)} \right|, \qquad \forall i, j, \tag{2.28}$$

and by Lemma 2.6 there exist $O \in O(n)$ and a vector $b \in \mathbb{R}^n$ such that $w^{(i)} = O v^{(i)} + b$, $i = 1, \ldots, n$. In terms of the corresponding matrices this means that $A^Y = A^X O^T + B$, where the matrix B is defined as $B_{i,j} := b_j$. We obtain therefore

$$Y = A^X O^T Z + B Z. \tag{2.29}$$

The random vector $A^X O^T Z$ is a centered Gaussian random vector with covariance $A^X O^T O (A^X)^T = C^X$ so that it has the same law of X. The random vector BZ has all the components equal and setting $W = \sum_{j=1}^n b_j Z_j$ we finish the proof. $\qquad \square$

A direct consequence of the above result is the following. Define the function $F : \bar{\mathcal{C}} \to \mathbb{R}$ by

$$F(C) := \mathbb{E}\left[\log \sum_{i=1}^n w_i e^{X_i} \right], \tag{2.30}$$

where X is a centered Gaussian random vector with covariance C.

Lemma 2.8 *Given $C^X, C^Y \in \bar{\mathcal{C}}$ such that $d_X = d_Y$, then $F(C^X) = F(C^Y)$.*

Proof Since $d_X = d_Y$ by Lemma 2.7 we have that $Y = X + W$ and therefore

$$F(C^Y) = \mathbb{E}\left[\log \sum_{i=1}^n w_i e^{Y_i} \right] = \mathbb{E}\left[\log \sum_{i=1}^n w_i e^{X_i + W} \right]$$
$$= \mathbb{E}\left[W + \log \sum_{i=1}^n w_i e^{X_i} \right] = F(C^X),$$

where the last equality follows by the fact that W is centered. $\qquad \square$

This Lemma simply says that we can define the right hand side of (2.30) as $\widetilde{F}(d)$ since the function depends just on the metric structure of the random variables and not on their correlations.

We expect therefore to have a version of Theorem 2.2 with conditions written just in terms of the metrics. This is done in the next section.

2.3 A Generalized Condition

We show how to generalize Theorem 2.2 proving that (2.17) can be deduced under weaker hypotheses concerning just the metric structures. The same inequality has been obtained in [11] with a tricky computation. Here we show that this fact follows from a general argument that may be applied for different functions f.

Theorem 2.9 *Let X, Y two centered Gaussian random vectors and the function f given by* (2.14). *If*

$$d_Y(i, j) \geq d_X(i, j) \qquad \forall i, j, \tag{2.31}$$

then

$$\mathbb{E}[f(Y)] \geq \mathbb{E}[f(X)]. \tag{2.32}$$

Note that if conditions (2.15) and (2.16) are satisfied then (2.31) holds, but it is easy to construct examples for which (2.31) holds but (2.15), (2.16) are violated.

Observe that for any x we have that $\mu(x) = (\mu_1(x), \ldots, \mu_n(x)) \in \mathcal{I}^n$ (recall definition (2.18)) where

$$\mathcal{I}^n = \left\{ \mu = (\mu_1, \ldots, \mu_n) : 0 \leq \mu_i \leq 1, \ \sum_{i=1}^n \mu_i = 1 \right\}.$$

Namely, $\mathcal{I}^n \subset \mathbb{R}^n$ is a $(n-1)$-dimensional simplex with extremal elements $\mu^{(1)}, \ldots, \mu^{(n)}$, where $\mu_i^{(l)} = \delta_{li}$.

We start with a preliminary Lemma

Lemma 2.10 *Consider a symmetric matrix D and the function $G : \mathcal{I}^n \to \mathbb{R}$ defined as*

$$G(\mu) := \sum_{i=1}^n \mu_i D_{ii} - \sum_{i=1}^n \sum_{j=1}^n \mu_i \mu_j D_{ij}. \tag{2.33}$$

We have that

$$\inf_{\mu \in \mathcal{I}^n} G(\mu) \geq 0 \tag{2.34}$$

if and only if

$$D_{ii} + D_{jj} - 2D_{ij} \geq 0 \qquad \forall i, j \in \{1, \ldots, n\}. \tag{2.35}$$

Proof If condition (2.35) holds, then

$$G(\mu) \geq \sum_{i=1}^n \mu_i D_{ii} - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \mu_i \mu_j (D_{ii} + D_{jj}) = 0.$$

To obtain the last identity we used the fact that $\mu \in \mathcal{I}^n$. Conversely, suppose inequality (2.34) to hold. Choose μ such that $\mu_l = \mu_m = \frac{1}{2}$ for some $l \neq m$ and 0 otherwise; then (2.33) becomes

$$\frac{1}{4}(D_{ll} + D_{mm}) - \frac{1}{2}D_{lm} \geq 0 \tag{2.36}$$

where we used the symmetry of D. Consider all the couples $l, m \in \{1, \ldots, n\}$ to get the result. □

Proof of Theorem 2.9 By formula (2.7) we deduce the results once we show that

$$\inf_{x \in \mathbb{R}^n} \left\{ \operatorname{Tr}\left(D \left(\nabla^2 f \right)_{i,j} (x) \right) \right\} \geq 0, \tag{2.37}$$

where we called

$$D := C^Y - C^X. \tag{2.38}$$

Using (2.19) and (2.20) we obtain that the expression to be minimized in (2.37) is

$$\sum_{i=1}^n D_{i,i} \mu_i (x) - \sum_{i=1}^n \sum_{j=1}^n D_{i,j} \mu_i (x) \mu_j (x). \tag{2.39}$$

We have therefore that the infimum in (2.37) coincides with $\inf_{\mu \in \mathcal{I}^n} G(\mu)$ and the result follows by Lemma 2.10 since (2.35) with the matrix D defined by (2.38) coincides with (2.31). □

3 Examples

In this section we discuss two examples, obtaining the existence of the thermodynamic limit for the quenched free energy of two models. The first one is the Sherrington–Kirkpatrick model. The existence of the thermodynamic limit for this model was obtained, by the interpolation method, in the breakthrough paper [19]. This was done using the result Theorem 2.2. We review this result as a warm-up to fix ideas and the basic constructions. We use however Theorem 2.9 and discuss the result just in terms of the metrics. Then we discuss a class of Generalized Random Energy Models [15] for which in general conditions (2.15), (2.16) fail while condition (2.31) hold. We refer to [9,25] and [10] for the beautiful mathematics involved in the limit of such kind of models. In the final part of the section we discuss some models from a purely metric viewpoint.

3.1 The Sherrington–Kirkpatrick Model

The Sherrington–Kirkpatrick model is a mean field spin glass model [18,24,26,28]. Spins configurations are $\sigma \in \{-1, 1\}^N$ and the energy of the system is given by

$$H_N (\sigma) := -\frac{1}{\sqrt{N}} \sum_{i,j=1}^N J_{i,j} \sigma (i) \sigma (j), \tag{3.1}$$

where $J_{i,j}$ are i.i.d. standard Gaussian random variables. Small variants of the model consider different sums in (3.1) but all the variants are equivalent modulo simple transformations. The spins are associated to the vertices of a complete graph and the interaction between each pair of spins is determined by the variables J's. The partition function is defined as

$$Z_N (\beta) := \sum_{\{\sigma\}} e^{-\beta H_N(\sigma)}, \tag{3.2}$$

where the parameter β is the inverse temperature and the quenched free energy per site is defined by

$$F_N (\beta) := -\frac{1}{\beta N} \mathbb{E} \left[\log Z_N (\beta) \right] := \frac{1}{\beta N} \alpha_N (\beta), \tag{3.3}$$

where the last equality defines the symbol $\alpha_N(\beta)$. The variables $(-\beta H_N(\sigma))_{\sigma \in \{-1,1\}^N}$ are a centered Gaussian random vector with covariance

$$\beta^2 \mathbb{E}\left[H_N(\sigma) H_N(\sigma')\right] = \frac{\beta^2}{N} \sum_{i,j=1}^{N} \sigma(i)\sigma(j)\sigma'(i)\sigma'(j) = N\beta^2 q_N^2(\sigma,\sigma'), \quad (3.4)$$

where

$$q_N(\sigma,\sigma') := \frac{1}{N} \sum_{i=1}^{N} \sigma(i)\sigma'(i), \quad (3.5)$$

is the *overlap* between the configurations σ and σ'. The corresponding Euclidean distance according to (2.23) is given by

$$d_N(\sigma,\sigma') = \beta\sqrt{8N\left[d_N^H\left(1 - d_N^H\right)\right]}, \quad (3.6)$$

where

$$d_N^H(\sigma,\sigma') := \frac{1}{N} \sum_{i=1}^{N} \mathbb{I}\left(\sigma(i) \neq \sigma'(i)\right) \quad (3.7)$$

is the *Hamming distance*. By the way, the Hamming distance is an example of a non Euclidean metric. Notice that we have of course $d_N(\sigma,\sigma) = 0$ but we have also $d_N(\sigma,-\sigma) = 0$ since $H_N(\sigma) = H_N(-\sigma)$. The fact that the right hand side of (3.7) is a distance (indeed a pseudo distance) is not trivial but follows directly since it is obtained by (2.23) (it is a function of a metric that is again a metric, see [13,16]).

Let us split the system into two subsystems S_1, S_2 with respectively N_1 and N_2 vertices with $N_1 + N_2 = N$. We erase the interaction between spins that belong to different subsystems. We define the restricted Hamiltonians of the subsystems as

$$H_{N_k}(\sigma) := -\frac{1}{\sqrt{N_k}} \sum_{i,j \in S_k} J_{i,j}\sigma(i)\sigma(j), \quad k = 1, 2, \quad (3.8)$$

where we remark that the sum is restricted to the indices belonging to the subsystems labeled $k = 1, 2$. Here and hereafter we continue to use the symbol σ both for the full configuration as well as for the configuration restricted to a subsystem. When a configuration appears in an expression that is labeled by a subsystem then we mean the configuration restricted to the subsystem. For example $d_{N_k}^H(\sigma,\sigma')$ and $d_{N_k}(\sigma,\sigma')$ are respectively the Hamming distance (3.7) and the distance (3.6) when the configuration is restricted to the subsystem $k = 1, 2$. Note that with this notation we have the key relationship

$$d_N^H(\sigma,\sigma') = \frac{N_1}{N} d_{N_1}^H(\sigma,\sigma') + \frac{N_2}{N} d_{N_2}^H(\sigma,\sigma'). \quad (3.9)$$

Another important relationship is

$$\sum_{\{\sigma\}} e^{-\beta\left(H_{N_1}(\sigma) + H_{N_2}(\sigma)\right)} = Z_{N_1}(\beta) Z_{N_2}(\beta). \quad (3.10)$$

We apply Theorem 2.9 with the vectors

$$\begin{cases} Y = \left(-\beta H_N(\sigma)\right)_{\sigma \in \{-1,1\}^N}, \\ X = \left(-\beta H_{N_1}(\sigma) - \beta H_{N_2}(\sigma)\right)_{\sigma \in \{-1,1\}^N}. \end{cases}$$

The condition (2.31) becomes the super-Pythagorean relation

$$d_N \geq \sqrt{d_{N_1}^2 + d_{N_2}^2},$$ (3.11)

that is equivalent to

$$\left[d_N^H \left(1 - d_N^H \right) \right] \geq \frac{N_1}{N} \left[d_{N_1}^H \left(1 - d_{N_1}^H \right) \right] + \frac{N_2}{N} \left[d_{N_2}^H \left(1 - d_{N_2}^H \right) \right].$$ (3.12)

The above inequality is true by (3.9) and the concavity of the real function $x \to x \, (1 - x)$. By Theorem 2.9 and (3.10) we deduce

$$\alpha_N \left(\beta \right) \leq \alpha_{N_1} \left(\beta \right) + \alpha_{N_2} \left(\beta \right),$$ (3.13)

and by sub-additivity and the classic Fekete Lemma we deduce that the limit of the quenched free energy per site exists

$$\lim_{N \to \infty} F_N \left(\beta \right) = \lim_{N \to \infty} \frac{1}{\beta N} \alpha_N \left(\beta \right) = \inf_N \frac{1}{\beta N} \alpha_N \left(\beta \right).$$ (3.14)

3.2 The Generalized Random Energy Model

The *Generalized Random Energy Model* (GREM) is a spin glass model introduced by Derrida [15] to generalize the REM (*Random Energy Model*) imposing pair correlations between energies. The model has a hierarchical structure, as any spin configuration correspond to a leaf of a given rooted tree.

We consider sequences of finite trees codified by finite strings of non-negative integers. Let $n \in \mathbb{N}$ and $\underline{k} = (k_1, \ldots, k_n)$ a vector of non-negative integers and call $|\underline{k}| := k_1 + \ldots + k_n$. The tree $\mathcal{T}_{\underline{k}}$ is constructed as follows. The root (that is the unique node at level 0) is connected to 2^{k_1} nodes to compose the first level. Each node of the first level is connected to 2^{k_2} nodes of the second level; we have therefore $2^{k_1 + k_2}$ nodes on the second level and so on. The n-th level consists of $2^{k_1} 2^{k_2} \ldots 2^{k_n} = 2^{|\underline{k}|}$ leaves. If there exists a $1 \leq j < n$ such that $k_j = 0$, we mean that the nodes of the level j coincide with those of the level $j - 1$. A spin configuration $\sigma \in \{-1, 1\}^{|\underline{k}|}$ is then attached to each leaf. The Hamiltonian is

$$H_{\underline{k}} \left(\sigma \right) = -\sqrt{|\underline{k}|} \left(\varepsilon_1^{(\sigma)} + \cdots + \varepsilon_n^{(\sigma)} \right),$$ (3.15)

where $\varepsilon_i^{(\sigma)} \sim \mathcal{N} \left(0, a_i \right)$ if $k_i > 0$ and $\varepsilon_i^{(\sigma)} = 0$ if $k_i = 0$. For any $i \in \mathbb{N}$ we have that the a_i's are positive numbers such that $\sum_{i=1}^{+\infty} a_i = 1$.

The random variables ε's are attached to the edges of the tree. More precisely attached to the edges that connect the level $i - 1$ to the level i there is a family of i.i.d. centered Gaussian random variables with variance a_i, one for each edge. When we write $\varepsilon_i^{(\sigma)}$ we mean then the random variable associated to the unique edge that connects level $i - 1$ to level i and that belongs to the unique path from the leaf associated to σ to the root. When $k_i = 0$ there are no edges from level $i - 1$ to level i and therefore we set $\varepsilon_i^{(\sigma)} = 0$. Then, $\left(H_{\underline{k}}(\sigma) \right)_{\sigma \in \{-1,1\}^{|\underline{k}|}}$ is a centered Gaussian random vector on the $|\underline{k}|$-dimensional hypercube $\{-1, 1\}^{|\underline{k}|}$.

We call $l = l \, (\sigma, \tau) \in \{0, 1, \ldots n - 1\}$ the level of the hierarchy at which the two paths from the leaves σ and τ of $\mathcal{T}_{\underline{k}}$ to the root merge. The two configurations share the same energy variables $\varepsilon_i^{(\sigma)} = \varepsilon_i^{(\tau)}$ for any $i \leq l$, while $\varepsilon_i^{(\sigma)} \neq \varepsilon_i^{(\tau)}$ whenever $i > l$. When $\varepsilon_i^{(\sigma)}$ and $\varepsilon_i^{(\tau)}$ are different, they are independent. Furthermore, $\varepsilon_i^{(\sigma)}$ and $\varepsilon_j^{(\tau)}$ are always independent
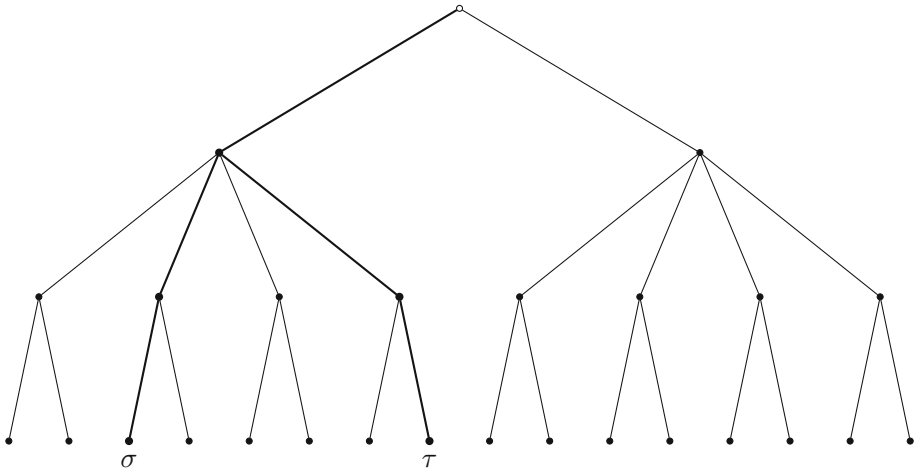
Fig. 2 The paths σ and τ are at distance $s_{\underline{k}}(\sigma, \tau) = 2(a_2 + a_3)$

if $j \neq i$. We define $\widetilde{a}_i := a_i$ when $k_i > 0$ and $\widetilde{a}_i := 0$ when $k_i = 0$. We get

$$\mathbb{E}\left[H_{\underline{k}}(\sigma) H_{\underline{k}}(\tau)\right] = |\underline{k}| \sum_{i=1}^{l} \widetilde{a}_i, \qquad (3.16)$$

pointing out that the right hand side above is zero when $l = 0$. The corresponding metric according to (2.23) is given by

$$d_{\underline{k}}(\sigma, \tau) = \sqrt{\mathbb{E}\left[\left(H_{\underline{k}}(\sigma) - H_{\underline{k}}(\tau)\right)^2\right]} = \sqrt{2|\underline{k}| \sum_{i=l+1}^{n} \widetilde{a}_i}. \qquad (3.17)$$

The term inside the square root on the right hand side represents, up to a multiplicative factor, the minimal path length distance between the two leaves σ and τ on the tree when each edge between level $i-1$ and i has a length given by a_i. Since the graph is a tree the path is unique and the metric (3.17) is an ultrametric. We introduce, for notational convenience, the normalized distance

$$s_{\underline{k}}(\sigma, \tau) := \sqrt{2 \sum_{i=l+1}^{n} \widetilde{a}_i}, \qquad (3.18)$$

so that $d_{\underline{k}}(\sigma, \tau) = \sqrt{|\underline{k}|} s_{\underline{k}}(\sigma, \tau)$ for any pair of configurations σ and τ.

Both the correlations (3.16) and the metric (3.17) depend on the vector $\underline{k}$ and on the assignment of configurations to leaves. We will discuss soon this.

Like for the Sherrington–Kirkpatrick model, given an inverse temperature β, we introduce the disorder-dependent partition function

$$Z_{\underline{k}}(\beta) := \sum_{\{\sigma\}} e^{-\beta H_{\underline{k}}(\sigma)} \qquad (3.19)$$

and the quenched average of the free energy per site

$$F_{\underline{k}}(\beta) := -\frac{1}{\beta|\underline{k}|} \mathbb{E}\left[\log Z_{\underline{k}}(\beta)\right]. \qquad (3.20)$$

We prove the existence of the thermodynamic limit of (3.20) under general assumptions when a parameter N is diverging and the vector $\underline{k} = \underline{k}(N)$ is growing in such a way that also $n = n(N)$ diverges. Contucci et al. [12] proved this fact when n is constant. This was obtained applying the same strategy of the Guerra–Toninelli interpolation method [19]; in particular, they used the inequality in Theorem 2.2. When n is no longer bounded this inequality fails while the inequality in Theorem 2.9 continues to work. We describe now more precisely the growing mechanism of the model and prove the existence of the thermodynamic limit.

3.2.1 Growing and Labeling

We consider a sequence of growing trees labeled by a sequence of vectors $\underline{k}(N)$. For each $N \in \mathbb{N}$ we have the tree $\mathcal{T}_{\underline{k}(N)}$ defined by the following hypothesis and rules.

(H1) Let $(\alpha_i)_{n=1}^{\infty}$ be a sequence of reals larger than 1 satisfying the constraint

$$\sum_{i=1}^{\infty} \log \alpha_i = \log 2. \tag{3.21}$$

The α_i's define the tree $\mathcal{T}_{\underline{k}(N)}$ through

$$k_i(N) := \left\lfloor \frac{N \log \alpha_i}{\log 2} \right\rfloor, \qquad i \in \mathbb{N}, \tag{3.22}$$

where $\lfloor \cdot \rfloor$ denotes the integer part.

(H2) The sequence $(a_i)_{i=1}^{\infty}$ corresponds to the lengths of the edges from the different levels and the variance of the associated random variables and satisfies the condition $\sum_{i=1}^{\infty} a_i = 1$.

The exact values of the sums of the series are not really important and could be substituted just by summability conditions. Formula (3.22) follows by the fact that we ask that the number of edges connecting a given node at level $i - 1$ to nodes at level i grows exponentially like α_i^N.

Observe that by (3.21), for any fixed $N > 0$ in $\underline{k}(N)$ just a finite number of components are different from zero. We define

$$n := n(N) := \max\{i \; : \; k_i(N) > 0\} \tag{3.23}$$

and the finite vector $\underline{k}(N) := (k_1(N), \ldots, k_n(N))$. Then, a spin configuration $\sigma \in \{-1, 1\}^{|\underline{k}(N)|}$ is assigned to each leaf. The method is actually arbitrary; indeed, the free energy of the system is obtained summing over all the configurations, thus getting rid of any dependence on the underlying choice.

We assign a spin configuration to each leaf of the tree as follows. At fixed N, we attach to every edge one or more labels of type (m, s), where $s = \pm 1$ and $m \in \{1, \ldots, |\underline{k}(N)|\}$. Given a leaf there exists a unique path toward the root. If this path crosses an edge having a label (m, s) then the configuration σ associated to the leave is such that $\sigma(m) = s$. We assign the labels in such a way that every path meets all the labels $m = 1, \ldots, |\underline{k}(N)|$ and such that different leaves have associated different configurations.

We embed the tree on a plane so that the root is on the top and the paths from the leaves to the root are going upwards. Moreover all the edges connecting a given node with the nodes at the successive level are ordered from left to right. Each edge connecting the level $i - 1$ to level i has exactly $k_i(N)$ labels corresponding to the values $m = \sum_{j=1}^{i-1} k_j(N) +$
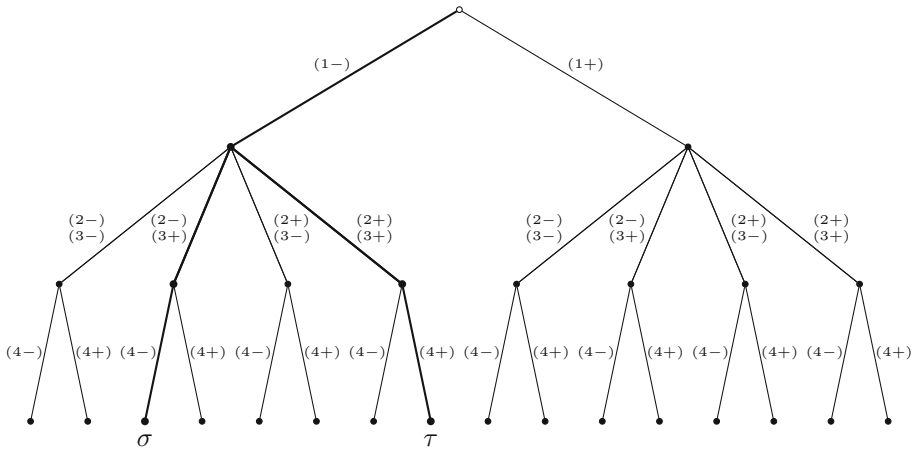
Fig. 3 Example of assignation of the labels for $\underline{k} = (1, 2, 1)$. Paths σ and τ have spin configurations $\sigma = (-1, -1, 1, -1)$, $\tau = (-1, 1, 1, 1)$

$1, \sum_{j=1}^{i-1} k_j(N) + 2, \ldots, \sum_{j=1}^{i} k_j(N)$. The corresponding values of the parameter s are fixed as follows.

Fix a node at level $i - 1$. Number each edge connecting this node with a node at level i with an integer number going from left to right from the value 0 to $2^{k_i(N)} - 1$. The leftmost will correspond to 0 while the rightmost to $2^{k_i(N)} - 1$. Do this for each node. Write these integers in binary code so that the leftmost edges are numbered with $k_i(N)$ zeros and the rightmost with $k_i(N)$ ones. In our setting, the 0 corresponds to the $-$ sign and the 1 to the $+$ sign. Then, we associate the lowest value of m to the most significant digit and the highest value of m to the less significant one. See Fig. 3 for an example.

3.2.2 Splitting the System

Let $N > 0$ and consider a pair of integers N_1, N_2 such that $N_1 + N_2 = N$. We already know how to construct the trees $\mathcal{T}_{\underline{k}(N)}$, $\mathcal{T}_{\underline{k}(N_1)}$ and $\mathcal{T}_{\underline{k}(N_2)}$. Their geometric structure is simply codified by the finite vectors $\underline{k}(N)$, $\underline{k}(N_1)$, $\underline{k}(N_2)$ and we recall that, by definition, we have

$$k_i(N_j) := \left\lfloor \frac{N_j \log \alpha_i}{\log 2} \right\rfloor, \qquad j = 1, 2, \quad i \in \mathbb{N}. \tag{3.24}$$

Notice that

$$k_i(N_1) + k_i(N_2) \leq k_i(N) \leq k_i(N_1) + k_i(N_2) + 1. \tag{3.25}$$

We associate the labels to the edges and leaves of the full system $\mathcal{T}_{\underline{k}(N)}$ as in the previous section. The labels of the two subsystems $\mathcal{T}_{\underline{k}(N_1)}$ and $\mathcal{T}_{\underline{k}(N_2)}$ are instead attributed in a slightly different way in order to have different spins (different labels m) belonging to the two subsystems.

The labels m attributed to the edges from level $i - 1$ to level i in the full system coincide with the set $\left\{ \sum_{j=1}^{i-1} k_j(N) + 1, \sum_{j=1}^{i-1} k_j(N) + 2, \ldots, \sum_{j=1}^{i} k_j(N) \right\}$. When we split the system into the two subsystems we assign to the edges that connect each node in the level $i - 1$ to the level i of the subsystem $\mathcal{T}_{\underline{k}(N_1)}$ the labels $\left\{ \sum_{j=1}^{i-1} k_j(N) + 1, \ldots, \sum_{j=1}^{i-1} k_j(N) + k_i(N_1) \right\}$

while we assign to the edges that connect each node in the level $i-1$ to the level i of the subsystem $\mathcal{T}_{\underline{k}(N_2)}$ the labels $\left\{\sum_{j=1}^{i-1} k_j(N) + k_i(N_1) + 1, \ldots, \sum_{j=1}^{i-1} k_j(N) + k_i(N_1) + k_i(N_2)\right\}$. By (3.25) this is well defined. Once split the labels m into the two subsystems, the assignment of the label $s = \pm$ follows the same rule of the previous section. Since $k_i(N_1) + k_i(N_2)$ may be strictly less than $k_i(N)$, some of the labels m (i.e. some spins) may disappear in the splitting.

We discuss now the behavior of the distances. Consider two finite vectors $\underline{k}$ and $\underline{k}'$ such that $k'_i \le k_i$ for any i. We assign the labels to $\mathcal{T}_{\underline{k}}$ in the usual way while instead we assign the labels to $\mathcal{T}_{\underline{k}'}$ as follows. We assign to the edges that connect each node in the level $i-1$ to the level i of $\mathcal{T}_{\underline{k}'}$ arbitrarily k'_i of the k_i labels in $\mathcal{T}_{\underline{k}}$. The assignment of the labels $s = \pm$ follows then the usual rule.

We call respectively $d_{\underline{k}}$ and $d_{\underline{k}'}$ the metrics defined by formula (3.17) for the two trees $\mathcal{T}_{\underline{k}}$ and $\mathcal{T}_{\underline{k}'}$ and $s_{\underline{k}}, s_{\underline{k}'}$ the corresponding normalized distances (see (3.18)). As before given two spin configurations $\sigma, \tau \in \{-1, 1\}^{|\underline{k}|}$ we call again $\sigma, \tau \in \{-1, 1\}^{|\underline{k}'|}$ the same configurations but restricted just to the labels assigned to the edges in $\mathcal{T}_{\underline{k}'}$. We have the following.

Lemma 3.1 *Consider two finite vectors $\underline{k}' \le \underline{k}$ and the corresponding trees $\mathcal{T}_{\underline{k}}$ and $\mathcal{T}_{\underline{k}'}$ with configurations of spins associated to the leaves as above. Then we have*

$$s_{\underline{k}'}(\sigma, \tau) \le s_{\underline{k}}(\sigma, \tau), \qquad \forall \sigma, \tau. \tag{3.26}$$

Proof Consider the tree $\mathcal{T}_{\underline{k}}$, two configurations σ, τ associated to two leaves and their corresponding geodetic path. Let us now consider a new finite vector $\underline{k}'$ obtained by $\underline{k}$ simply decreasing by one just a single component and preserving all the remaining ones, i.e. $k'_i = k_i - 1$ and $k'_j = k_j$ for all $j \ne i$. Suppose that the label m that is missing in $\mathcal{T}_{\underline{k}'}$ is m^*. The tree $\mathcal{T}_{\underline{k}'}$ with the corresponding labeling is obtained from $\mathcal{T}_{\underline{k}}$ and the original labeling simply as follows. All the edges connecting nodes at level $i-1$ to nodes at level i in $\mathcal{T}_{\underline{k}}$ can be paired into pairs having exactly the same labels apart the one corresponding to m^*. The two paired edges will have labels respectively $(m^*, +)$ and $(m^*, -)$. If we identify each paired couple of edges, and consequently we identify too the subtrees starting from the identified nodes, we get a tree that coincides with $\mathcal{T}_{\underline{k}'}$ with exactly the same assignments of labels. In particular, the leaves associated to σ, τ in the new tree will be exactly the original ones after the identification. Finally the geodetic path too remains the same after the identification (see e.g. Fig. 4).

Since the identification procedure can only shorten this path we have the statement of the lemma when $\underline{k}'$ is obtained by $\underline{k}$ decreasing by one just one of its components. We finish the proof observing that any $\underline{k}' \le \underline{k}$ can be obtained by $\underline{k}$ after a finite numbers of iterations of this type. $\square$

Remark 3.2 Both $\mathcal{T}_{\underline{k}(N_i)}$, $i = 1, 2$ are obtained by $\mathcal{T}_{\underline{k}(N)}$ as in the hypothesis of Lemma 3.1 and we have therefore

$$s_{\underline{k}(N)}(\sigma, \tau) \ge \max\left\{s_{\underline{k}(N_1)}(\sigma, \tau), s_{\underline{k}(N_2)}(\sigma, \tau)\right\}, \qquad \forall \sigma, \tau, \quad i = 1, 2. \tag{3.27}$$

Since by (3.25) we have $\frac{|\underline{k}(N_1)| + |\underline{k}(N_2)|}{|\underline{k}(N)|} \le 1$ and we deduce

$$s_{\underline{k}(N)}^2(\sigma, \tau) \ge \frac{|\underline{k}(N_1)|}{|\underline{k}(N)|} s_{\underline{k}(N_1)}^2(\sigma, \tau) + \frac{|\underline{k}(N_2)|}{|\underline{k}(N)|} s_{\underline{k}(N_2)}^2(\sigma, \tau), \tag{3.28}$$
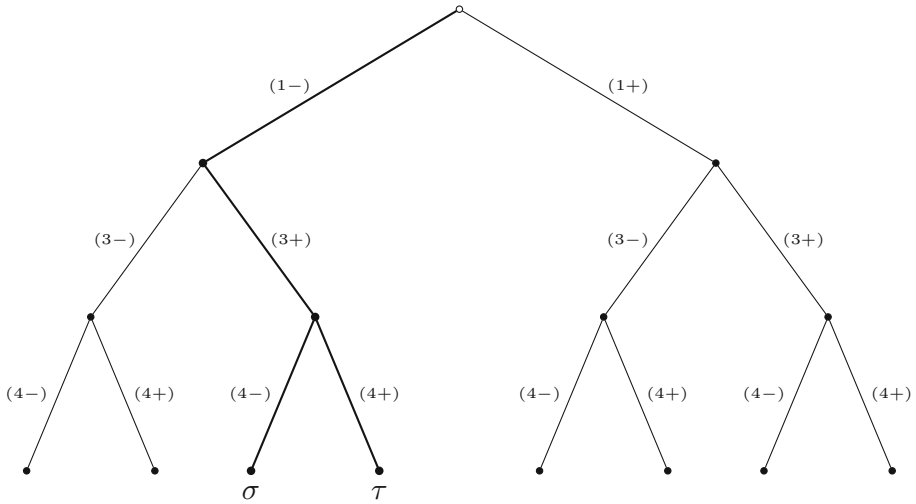
Fig. 4 After the coalescence of the branches with $m^* = 2$, the configurations σ and τ in Figs. 3 and 4 are at distance $s_{\underline{k}'}(\sigma, \tau) = \sqrt{2a_3} < s_{\underline{k}}(\sigma, \tau) = \sqrt{2(a_2 + a_3)}$

that is equivalent to the super-Pythagorean condition

$$d_{\underline{k}(N)}(\sigma, \tau) \geq \sqrt{d^2_{\underline{k}(N_1)}(\sigma, \tau) + d^2_{\underline{k}(N_2)}(\sigma, \tau)}. \tag{3.29}$$

3.2.3 Thermodynamic Limit

We define the energy of our sequence of GREM models as $H_N(\sigma) := H_{\underline{k}(N)}(\sigma)$ (recall definition (3.15)) and the corresponding partition functions and density of free energies like in (3.19), (3.20) more precisely $Z_N(\beta) = \sum_{\{\sigma\}} e^{-\beta H_N(\sigma)}$ and

$$F_N(\beta) = -\frac{1}{\beta |\underline{k}(N)|} \mathbb{E}\left[\log Z_N(\beta)\right] =: \frac{\alpha_N(\beta)}{\beta |\underline{k}(N)|}, \tag{3.30}$$

where the last equality defines the symbol $\alpha_N(\beta)$.

We need a preliminary Lemma. Let us call $\gamma_i := \frac{\log \alpha_i}{\log 2} > 0$. Observe that by definition we have $\sum_{i=1}^{+\infty} \gamma_i = 1$.

Lemma 3.3 *We have*

$$\lim_{N \to \infty} \frac{|\underline{k}(N)|}{N} = 1 \tag{3.31}$$

Proof For any finite k we have

$$\frac{\sum_{i=1}^{k}(N\gamma_i - 1)}{N} \leq \frac{|\underline{k}(N)|}{N} \leq \frac{\sum_{i=1}^{+\infty} N\gamma_i}{N}. \tag{3.32}$$

The right hand side of the above equation is 1. The left hand side converges when $N \to \infty$ to $\sum_{i=1}^{k} \gamma_i$. Taking now the limit on $k \to \infty$ we deduce the statement of the Lemma. □

We can now prove the existence of the limit for quenched free energy per site of a GREM model with infinite levels.

Theorem 3.4 *Under the hypothesis (H1) and (H2), there exists the limit when $N \to \infty$ of the density of free energy (3.30) defined on $T_{\underline{k}(N)}$, in the sense that there exists the following limit that coincides with an infimum*

$$-\infty < \lim_{N \to \infty} -\frac{1}{\beta \left| \underline{k}(N) \right|} \mathbb{E}\left[\log Z_N(\beta) \right] = \inf_N \frac{\alpha_N(\beta)}{\beta N} < \infty. \tag{3.33}$$

Proof We apply the interpolation method for the Gaussian random vectors $H_{\underline{k}(N)}(\sigma)$ and $H_{\underline{k}(N_1)}(\sigma) + H_{\underline{k}(N_2)}(\sigma)$ that are both labeled by the configurations $\sigma \in \{-1, 1\}^{\left| \underline{k}(N) \right|}$. The Gaussian random variables used to compute $H_{\underline{k}(N)}$, $H_{\underline{k}(N_1)}$ and $H_{\underline{k}(N_2)}$ are all independent among them. Note that since in the splitting some spins are lost then the second Gaussian random vector is degenerate.

We have the following identity

$$\sum_{\{\sigma\}} e^{-\beta H_{\underline{k}(N_1)}(\sigma)} e^{-\beta H_{\underline{k}(N_2)}(\sigma)} = Z_{N_1}(\beta) Z_{N_2}(\beta) 2^{\left| \underline{k}(N) \right| - \left| \underline{k}(N_1) \right| - \left| \underline{k}(N_2) \right|}. \tag{3.34}$$

The last term is due to the fact that some spins may be lost in the splitting.

By Remark 3.2, we can apply Theorem (2.9) getting

$$\alpha_N(\beta) \leq \alpha_{N_1}(\beta) + \alpha_{N_2}(\beta) - \left(\left| \underline{k}(N) \right| - \left| \underline{k}(N_1) \right| - \left| \underline{k}(N_2) \right| \right) \log 2. \tag{3.35}$$

Since the last term in the above inequality is non-negative we obtain that the sequence $\alpha_N(\beta)$ is subadditive. By Fekete's Lemma we deduce that there exists the limit

$$\lim_{N \to \infty} \frac{\alpha_N(\beta)}{\beta N} = \inf_N \frac{\alpha_N(\beta)}{\beta N}. \tag{3.36}$$

By Lemma 3.3 we have that

$$\lim_{N \to \infty} \frac{\alpha_N(\beta)}{\beta \left| \underline{k}(N) \right|} = \lim_{N \to \infty} \frac{\alpha_N(\beta)}{\beta N}, \tag{3.37}$$

and we get the main statement of the Theorem.

It remains just to prove that the limit is strictly bigger than $-\infty$.

This follows by the summability of the variances a_i's. Indeed, we prove that for any $N > 0$, $-\beta F_N(\beta)$ is bounded from above. We have

$$-\beta F_N(\beta) = \frac{1}{\left| \underline{k}(N) \right|} \mathbb{E}\left[\log Z_N(\beta) \right] = \frac{1}{\left| \underline{k}(N) \right|} \mathbb{E}\left[\log \sum_{\{\sigma\}} e^{\beta(\varepsilon_1^{(\sigma)} + \varepsilon_2^{(\sigma)} + \ldots + \varepsilon_{n(N)}^{(\sigma)})} \right]$$

$$\leq \frac{1}{\left| \underline{k}(N) \right|} \log \sum_{\{\sigma\}} \mathbb{E}\left[e^{\beta(\varepsilon_1^{(\sigma)} + \varepsilon_2^{(\sigma)} + \ldots + \varepsilon_{n(N)}^{(\sigma)})} \right], \tag{3.38}$$

where we used Jensen's inequality. Since the $\varepsilon_i^{(\sigma)}$ are independent, the expectation value in the last row is the product of generating functions:

$$\mathbb{E}\left[e^{\beta \varepsilon_i^{(\sigma)}} \right] = e^{\frac{\beta}{2} a_i} \qquad \forall i, \tag{3.39}$$

hence

$$-\beta F_N(\beta) \leq \frac{1}{\left| \underline{k}(N) \right|} \log \sum_{\{\sigma\}} e^{\frac{\beta}{2} \sum_{i=1}^{n(N)} a_i} \leq \frac{1}{\left| \underline{k}(N) \right|} \log \sum_{\{\sigma\}} e^{\frac{\beta}{2}} = \log 2 + \frac{\beta}{2 \left| \underline{k}(N) \right|}, \tag{3.40}$$

where we used the fact that $\sum_{i=1}^{\infty} a_i = 1$.   □

Just as a remark we show in Lemma 3.5 in the Appendix that the third term in the right hand side of (3.35) is negligible when N is large. This fact is irrelevant for the proof but it is interesting in itself since for different models we could have a similar situation but with the wrong sign and a bound of this type could allow to apply the generalized subadditive lemmas in [14].

3.3 Geometric Remarks

Since by Lemma 2.8 we have that the density of free energy just depends on the metric structure of the Gaussian random variables, it is interesting to analyze the metric structure corresponding to the different models. Moreover natural and interesting models can be introduced starting directly from the metric description. Since all the metric spaces involved must be Euclidean a relevant characteristic is the dimension of the space where the metric can be realized as a collection of points.

Another useful remark is that the super-Pythagorean relation (3.11) implies (3.13), that gives the convergence (3.14) to the infimum of the density of free energies while instead a sub-Pythagorean relation (i.e. (3.11) with the opposite inequality) would imply a convergence of the density of free energy to the supremum.

Let us start with the Sherrington–Kirkpatrick model. Since the energy is defined in terms of N^2 i.i.d. Gaussian random variables the metric of the model with N sites can be represented by 2^N points embedded on $\mathbb{R}^{N^2}$. Indeed a natural representation of this metric is the following. Consider $\sigma \in \{-1, 1\}^N$ as a column vector and define the projector $\widehat{\sigma} := \sqrt{N} \sigma \sigma^T$ that is a positive definite $N \times N$ matrix. By a direct computation we have that the metric induced by the Sherrington–Kirkpatrick model is given by

$$d_N(\sigma, \eta) = \sqrt{\mathrm{Tr}\left((\widehat{\sigma} - \widehat{\eta})^2\right)},$$

i.e. it is the Euclidean metric on the one dimensional projectors induced by the Hilbert-Schmidt scalar product. The super-Pythagorean relation is strictly related to the fact that all the projectors belong to the cone of positive definite matrices.

The metric structure of the GREM is best described with the trees illustrated in Sect. 3.2.

An interesting class of models can be introduced directly defining the metric. Let $v^{\pm 1} \in \mathbb{R}^2$ be two vectors. To any $\sigma \in \{-1, 1\}^N$ we associate the vector $v(\sigma) \in \mathbb{R}^{2^N}$ defined by

$$v(\sigma) := \otimes_{i=1}^N v^{\sigma(i)}. \tag{3.41}$$

We have

$$\left(v(\sigma), v(\eta)\right) = \prod_{i=1}^N \left(v^{\sigma(i)}, v^{\eta(i)}\right).$$

If we let $v^{\pm 1} = v^{\pm 1, N}$ depend on N in such a way that $|v^{\pm 1, N}|^2 = N^{\frac{1}{N}}$ and $\left(v^{1, N}, v^{-1, N}\right) = N^{\frac{1}{N}} \alpha^{\frac{1}{N}}$, $\alpha \in [0, 1)$ we obtain that the Euclidean metric on the 2^N points embedded into $\mathbb{R}^{2^N}$ (that are the points in correspondence of the vectors) is given by

$$d_N(\sigma, \eta) = \sqrt{2N \left(1 - \alpha^{d_N^H(\sigma, \eta)}\right)}. \tag{3.42}$$

This Euclidean metric (as before it is a function of a metric that is again a metric [13,16]) satisfies the super-Pythagorean relation (3.11) since the real function $1 - \alpha^x$ is concave. We have therefore convergence of the free energy densities. A model that corresponds to this metric can be fixed such that $\mathbb{E}\left[H_N(\sigma) H_N(\eta)\right] \sim \alpha^{d_N^H(\sigma,\eta)}$. The special case $\alpha = 0$ corresponds to the REM model that is a special case of the GREM discussed in Sect. 3.2 with one single branch and all the leaves directly connected to the root. In this case all the points are equally spaced.

Appendix

We show here that the extra terms in (3.35) are indeed negligible when N is large.

Lemma 3.5 *We have*

$$\lim_{N \to \infty} \sup_{N_1 + N_2 = N} \frac{\left|\underline{k}(N)\right| - \left|\underline{k}(N_1)\right| - \left|\underline{k}(N_2)\right|}{\left|\underline{k}(N)\right|} = 0. \tag{3.43}$$

Proof Let us call $I(N) \subseteq \mathbb{N}$ the set $I(N) := \{i : k_i(N) > 0\}$. We define also $J(N) := I(N) \cap (I(N-1))^C$. We have for any $N_1 + N_2 = N$ that

$$\left|\underline{k}(N)\right| - \left|\underline{k}(N_1)\right| - \left|\underline{k}(N_2)\right| \le |I(N)| \tag{3.44}$$

By definition we have that if $i \in J(N)$ then $\frac{1}{N} \le \gamma_i < \frac{1}{N-1}$. We have therefore

$$1 = \sum_{i=1}^{+\infty} \gamma_i \ge \sum_{\ell=1}^{+\infty} |J(\ell)| \frac{1}{\ell}. \tag{3.45}$$

We deduce therefore that the series on the right hand side has to be convergent. Since $|I(N)| = \sum_{\ell=1}^{N} |J(\ell)|$ we have

$$\frac{|I(N)|}{N} = \sum_{\ell=1}^{N} \frac{\ell}{N} |J(\ell)| \frac{1}{\ell}. \tag{3.46}$$

The series on the right hand side of (3.45) is convergent, thus we deduce the statement by the dominated convergence Theorem. This, together with (3.44) and Lemma 3.3, concludes the proof.   □

References

1. Alaoui, A.E., Krzakala, F., Vail, C.O.: Estimation in the spiked wigner model: a short proof of the replica formula. IEEE Int. Symp. Inf. Theory **2018**, 1874–1878 (2018)
2. Barbier, J., Dia, M., Macris, N., Krzakala, F.: The mutual information in random linear estimation 54th Annual Allerton Conference on Communication, Control, and Computing (Allerton), pp. 625–632. IL, Monticello (2016)
3. Barbier, J., Macris, N.: The adaptive interpolation method: a simple scheme to prove replica formulas in Bayesian inference Probab. Theory Relat. Fields **174**, 1133–1185 (2019)
4. Barbier, J., Macris, N., Dia, M., Krzakala, F.: Mutual information and optimality of approximate message-passing in random linear estimation. IEEE Trans. Inf. Theory **66**(7), 4270–4303 (2020)
5. Barra, A.: The mean field Ising model through interpolating techniques. J. Stat. Phys. **132**(5), 787–809 (2008)
6. Barra, A., Contucci, P., Mingione, E., Tantari, D.: Multi-species mean field spin glasses. Rigorous results. Ann. Henri Poincaré **16**(3), 691–708 (2015)
7. Bhatia, R.: Positive definite matrices Princeton Series in Applied Mathematics. Princeton University Press, Princeton, NJ (2007)
8. Blumenthal, L.M.: Theory and Applications of Distance Geometry, 2nd edn. Chelsea Publishing Co., New York (1970)
9. Bovier, A., Kurkova, I.: Derrida's generalized random energy models. 1. Models with infinitely many hierarchies. Ann. Inst. H. Poincaré. Prob. Stat. **40**, 439–480 (2004)
10. Bovier, A., Kurkova, I.: Derrida's generalized random energy models. II. Models with continuous hierarchies. Ann. Inst. H. Poincaré Probab. Statist. **40**(4), 481–495 (2004)
11. Chatterjee, S.: An error bound in the Sudakov-Fernique inequality. arXiv:math/0510424
12. Contucci, P., Degli, Esposti M., Giardinà, C., Graffi, S.: Thermodynamical limit for correlated Gaussian random energy models. Commun. Math. Phys. **236**(1), 55–63 (2003)
13. Corazza, P.: Introduction to metric-preserving functions Amer. Math. Monthly **106**(4), 309–323 (1999)
14. de Bruijn, N.G. , Erdös, P.: Some linear and some quadratic recursion formulas. II. Nederl. Akad. Wetensch. Proc. Ser. A. **55** = Indagationes Math. **14**, 152–163 (1952)
15. Derrida, B.: A generalization of the random energy model that includes correlations between the energies. J. Phys. Lett. **46**, 401–407 (1985)
16. Doboš, J.: Metric Preserving Functions, Vydavatel'stvo Troffek, Koice, Slovakia, ISBN 80-88896-30-4 (1998)
17. Dokmanic, I., Parhizkar, R., Ranieri, J., Vetterli, M.: Euclidean Distance Matrices: Essential theory, algorithms, and applications. IEEE Signal Processing Magazine **32**, 6 (2015)
18. Guerra, F.: Spin Glasses. In: Bovier, A., et al. (eds.) Mathematical Statistical Physics, pp. 243–271. Elsevier, Oxford (2006)
19. Guerra, F., Toninelli, F.L.: The thermodynamic limit in mean field spin glass models. Commun. Math. Phys. **230**(1), 71–79 (2002)
20. Joag-dev, K., Perlman, M.D., Pitt, L.D.: Association of normal random variables and Slepian's inequality. Ann. Probab. **11**, 451–455 (1983)
21. Kahane, J.P.: Une inégalité du type Slepian and Gordon sur les processus gaussiens Israel. J. Math. **55**, 109–110 (1986)
22. Korada, S.B., Macris, N.: Exact solution of the gauge symmetric p-spin glass model on a complete graph. J. Stat. Phys. **136**(2), 205–230 (2009)
23. Mezard, M., Parisi, G., Virasoro, M.A.: Spin Glass Theory and Beyond. World Scientific, Singapore (1987)
24. Panchenko, D.: The Sherrington–Kirkpatrick Model Springer Monographs in Mathematics. Springer, New York (2013)
25. Ruelle, D.: A mathematical reformulation of Derrida's REM and GREM Comm. Math. Phys. **108**(2), 225–239 (1987)
26. Sherrington, D., Kirkpatrick, S.: Solvable model of a spin-glass. Phys. Rev. Lett. **35**, 1792–1796 (1975)
27. Schoenberg, I.: Remarks to Maurice Fréchets article Sur la définition axiomatique dune classe despace distanciés vectoriellement applicable sur lespace de Hilbert. Ann. Math. **36**, 724–732 (1935)
28. Talagrand M.: Spin glasses: a challenge for mathematicians. Cavity and mean field models. Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics, **46** Springer-Verlag, Berlin (2003)

29. Talagrand, M.: The Parisi formula. Ann. Math. (2) **163**(1), 221–263 (2006)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.