Chapter 1

Variational methods

(*Lesson 21 of 27/04/20*) *Compiled: May 18, 2020*

Exactly solvable models are rare. For example, the Ising Model, describing in a very simplified manner a discrete set of local interacting binary variables, has been exactly solved only for $d = 1$ in general, and for $d = 2$ only in absence of an external field ($h = 0$). The latter, in particular, requires long and sophisticated derivations.
Even for other models, the trend is the same: whenever we wish to study *emergent phenomena* the problem usually becomes analytically intractable.

One possibility is then to resort to **numerical simulations**. However, these are often time-consuming, require significant computational power, and can be hard to interpret - as interesting "high level" characteristics (such as the conditions for phase transitions) are drowned in lots of irrelevant "low-level" data.

So we may resort to **approximate computations** instead. The idea is to find a simple model that is able to capture, at least *qualitatively*, features from a more complex one, while still admitting an exact solution. This can then give hints on *what to look for* in a full numerical simulation, thus allowing a deeper understanding.

One quick way to compute approximations is through **variational methods**. In essence, we consider some parametrized pdf $f_{\boldsymbol{\theta}}(\boldsymbol{x})$, and tweak the parameters $\boldsymbol{\theta}$ so that it becomes "closer and closer" to the target pdf $f(\boldsymbol{x})$ of the full model. If we choose a sufficiently *simple* form for $f_{\boldsymbol{\theta}}$, we will be able to perform exact computations, while still retaining some sort of "correspondance" with the more complex model.

In the following, we will first introduce a notion of "**distance**" between pdf*s* (**relative entropy**), giving a mathematical meaning to the notion of "closeness" between probability distributions. Then we will explicitly state the *variational method* as a **minimization problem**, and, using the Ising Model as an example, we will see a popular choice for the parametrization of $f_{\boldsymbol{\theta}}$: the **mean-field approximation**.

1.0.1 Relative Entropy

Given two (discrete) probability distributions $\{p_i\}_{i\in\mathcal{D}}$ and $\{q_i\}_{i\in\mathcal{D}}$, with $p_i, q_i > 0$ and $\sum_i p_i = \sum_i q_i = 1$, we define the **relative entropy** (or Kullback–Leibler divergence) of $\{p_i\}$ with respect to $\{q_i\}$ as follows:

eqn:relative-entropy

$$S_R(\{p_i\},\{q_i\}) = -\sum_{i\in\mathcal{D}} p_i \ln \frac{p_i}{q_i} \leq 0 \tag{1.1}$$

In a sense, relative entropy measures the *closeness* between the two distributions - as it is maximum ($S_R = 0$) when the two coincide, i.e. $p_i = q_i\ \forall i$. Note, however, that S_R is not a *distance function* in the proper sense, as it does not satisfy the triangular inequality.

Proof that $S_R \leq 0$

The fact that $S_R = 0$ is the maximum point of S_R, i.e. $S_R \leq 0$, can be proven as follows. First we define an auxiliary function $f(x)$ over $(0,\infty)$:

$$f(x) = -x\ln x \qquad x > 0$$

Such function $f(x)$ is **concave**. In fact:

$$\begin{aligned} f'(x) &= -1 - \ln x \\ f''(x) &= -\frac{1}{x} < 0 \qquad x > 0 \end{aligned}$$

So, we may apply Jensen's inequality. For any choice of a set of non-negative numbers $\{\lambda_i\}$ summing to 1, the following relation holds:

$$f\left(\sum_i \lambda_i x_i\right) \geq \sum_i f(x_i)\lambda_i \qquad \sum_i \lambda_i = 1 \wedge \lambda_i \geq 0$$

And letting $\lambda_i = q_i$ and $x_i = p_i/q_i$ completes the proof:

$$S_R = \sum_i q_i f\left(\frac{p_i}{q_i}\right) \leq f\left(\sum_i \cancel{q_i}\frac{p_i}{\cancel{q_i}}\right) = f(1) = 0$$

with the equality holding if and only if $p_i = q_i$.

1.0.2 Approximation as an optimization problem

Let's consider, for simplicity, a system with **discrete** states $\{\boldsymbol{\sigma_i}\}_{i\in\mathcal{D}}$, each with energy $\mathcal{H}(\boldsymbol{\sigma_i})$, and an associated probability q_i given by a Boltzmann distribution:

$$\rho(\boldsymbol{\sigma_i}) \equiv q_i = \frac{e^{-\beta\mathcal{H}(\boldsymbol{\sigma_i})}}{Z} = e^{-\beta(\mathcal{H}(\boldsymbol{\sigma})-F)} \qquad Z = \sum_{\{\boldsymbol{\sigma}\}} e^{-\beta\mathcal{H}(\boldsymbol{\sigma})} \equiv e^{-\beta F}$$

where F is the system's **free energy** function.

In general, the $\{q_i\}$ are difficult to explicitly compute, because Z is generally a sum over a huge number of terms (2^V in the case of the Ising Model) with no analytical form.

So, the idea is to approximate ρ with another "easier" distribution ρ_0, the **variational ansatz**, which is parametrized as a Boltzmann distribution with a different Hamiltonian $\mathcal{H}_0$ (and so also a different free energy F_0):

$$\rho_0(\boldsymbol{\sigma_i}) \equiv p_i = \frac{e^{-\beta\mathcal{H}_0(\boldsymbol{\sigma_i})}}{Z_0} = e^{-\beta(\mathcal{H}_0(\boldsymbol{\sigma})-F_0)} \qquad Z_0 = \sum_{\{\boldsymbol{\sigma}\}} e^{-\beta\mathcal{H}_0(\boldsymbol{\sigma})} \equiv e^{-\beta F_0} \tag{1.2}$$

eqn:variational-ansatz

The *closeness* of $\{p_i\}$ to $\{q_i\}$ is given by their **relative entropy** (**??**):

$$\begin{aligned}
0 \leq \sum_i p_i \ln\frac{p_i}{q_i} &= \sum_{\{\boldsymbol{\sigma}\}} \frac{e^{-\beta\mathcal{H}_0(\boldsymbol{\sigma})}}{Z_0} \ln \frac{e^{-\beta\mathcal{H}_0(\boldsymbol{\sigma})}}{\underbrace{Z_0}_{e^{-\beta F_0}}} \frac{\overbrace{Z}^{e^{-\beta F}}}{e^{-\beta\mathcal{H}(\boldsymbol{\sigma})}} = \\
&= \frac{1}{Z_0}\sum_{\{\boldsymbol{\sigma}\}} e^{-\beta H_0(\boldsymbol{\sigma})}\beta[\mathcal{H}(\boldsymbol{\sigma}) - \mathcal{H}_0(\boldsymbol{\sigma}) - F + F_0] = \\
&= \beta\langle \mathcal{H} - \mathcal{H}_0\rangle_0 - \beta(F - F_0)
\end{aligned} \tag{1.3}$$

eqn:rel-entr

where $\langle\cdots\rangle_0$ denotes the average according to the ansatz distribution:

$$\langle f(\boldsymbol{\sigma})\rangle_0 \equiv \frac{1}{Z_0}\sum_{\{\boldsymbol{\sigma}\}} e^{-\beta\mathcal{H}_0(\sigma)} f(\boldsymbol{\sigma})$$

The expression (**??**) is called the **Gibbs-Bogoliubov-Feynman inequality**[1] bhfn:1, and holds as an equality if and only if $\rho = \rho_0 \Leftrightarrow \mathcal{H} = \mathcal{H}_0$.

Rearranging (**??**):

$$\beta F \leq \beta F_0 + \beta\langle\mathcal{H} - \mathcal{H}_0\rangle_0 = \beta\langle\mathcal{H}\rangle_0 + \beta(F_0 - \langle\mathcal{H}_0\rangle_0) \tag{1.4}$$

eqn:ineq-1

Note that F_0 does not depend on $\boldsymbol{\sigma}$, as it's $\propto \ln Z_0$, and so we can bring it inside the average, and expand it:

$$\beta(F_0 - \langle\mathcal{H}_0\rangle_0) = \beta\langle F_0 - \mathcal{H}_0\rangle_0 = \sum_{\{\boldsymbol{\sigma}\}} \rho_0(\boldsymbol{\sigma})\, \beta(F_0 - \mathcal{H}_0(\boldsymbol{\sigma}))$$

Then, from (**??**) note that:

$$\rho_0(\boldsymbol{\sigma}) = e^{-\beta(\mathcal{H}_0(\boldsymbol{\sigma})-F_0)} \Rightarrow \ln\rho_0(\boldsymbol{\sigma}) = \beta(F_0 - \mathcal{H}_0(\boldsymbol{\sigma}))$$

and substituting above:

$$\beta(F_0 - \langle\mathcal{H}_0\rangle_0) = -\frac{1}{k_B}\underbrace{\left(-k_B\sum_{\{\boldsymbol{\sigma}\}}\rho_0(\boldsymbol{\sigma})\ln\rho_0(\boldsymbol{\sigma})\right)}_{S[\rho_0]} = -\frac{S[\rho_0]}{k_B} \tag{1.5}$$

eqn:s-entropy

where $S[\rho_0]$ is the **information entropy** of ρ_0:

$$S[\rho_0] = -k_B\sum_{\{\boldsymbol{\sigma}\}}\rho_0(\boldsymbol{\sigma})\ln\rho_0(\boldsymbol{\sigma})$$

[1] ∧Physically, it is completely equivalent to the second law of thermodynamics.

Thus, substituting (**??**) back in the inequality (**??**) leads to:

$$\beta F \leq \beta \langle \mathcal{H} \rangle_0 - \frac{S[\rho_0]}{k_B} = \beta \langle \mathcal{H} \rangle_0 - \beta T S[\rho_0] \quad \texttt{eqn:var-principle} \tag{1.6}$$

And dividing by β:

$$F \leq F_V \equiv \langle \mathcal{H} \rangle_0 - T S[\rho_0]$$

where F_V is called the **Variational Free Energy** (VFE).
So, the true free energy F is always less or equal to the variational one F_V. An optimal estimate of F is obtained by minimizing F_V with respect to ρ_0.

Clearly, if we do not require any constraint on ρ_0, thus allowing arbitrary complexity, then the minimum is obtained when $\rho_0 = \rho$: the most accurate approximation of a model is the model itself. Realistically ρ is mathematically intractable, and we need to *bound* the "complexity" of ρ_0, with the effect that it won't be able to perfectly replicate ρ, and so the minimum for F_V will be larger than F (but hopefully still somewhat close).

One possible way to constrain the "complexity" of ρ_0 is to *force it* to be separable:

$$\rho_0(\boldsymbol{\sigma}) = \prod_x \rho_x(\sigma_x) \quad \texttt{eqn:mean-field} \tag{1.7}$$

In this way, all degrees of freedom of the system become **decoupled**. In a sense, correlations and complex behaviours are "averaged" between each component - and in fact the approximation in (**??**) is known as the **mean field** ansatz.

1.1 Mean Field Ising Model

Consider a d-dimensional nearest-neighbour Ising Model, where we allow each spin to interact with a **local** magnetic field b_x, leading to the Hamiltonian:

$$\mathcal{H}(\boldsymbol{\sigma}) = -J \sum_{\langle x,y \rangle} \sigma_x \sigma_y - \sum_x b_x \sigma_x$$

To understand its behaviour, we use the **mean-field** approximation (**??**), and choose a parametrization inspired by the non-interacting Ising Model (**??**, pag. **??**):

$$\rho_0(\boldsymbol{\sigma}) = \prod_x \rho_x(\sigma_x) \qquad \rho_x(\sigma_x) = \frac{1 + m_x \sigma_x}{2} \quad m_x \in [-1, 1] \quad \texttt{eqn:mfi} \tag{1.8}$$

where the $\{m_x\}$ are the *variational parameters* that will be *tweaked* to make $\rho_0(\boldsymbol{\sigma})$ closer to the real probability distribution $\rho(\boldsymbol{\sigma})$ of the Ising Model, by minimizing the **variational free energy** F_V. The constraint $m_x \in [-1, 1]$ comes from requiring all probabilities to be non-negative $\rho_x(\sigma_x) \geq 0$.
Before proceeding, note that (**??**) is already normalized:

$$\sum_{\sigma_x = \pm 1} \rho_x(\sigma_x) = \frac{1 + m_x}{2} + \frac{1 - m_x}{2} = \frac{1}{2} + \frac{1}{2} = 1$$

and that each *variational parameter* m_x corresponds to the **local magnetization** of spin σ_x *in the mean-field model*:

$$\begin{aligned}
\langle \sigma_x \rangle_0 &= \sum_{\{\boldsymbol{\sigma}\}} \rho_0(\boldsymbol{\sigma}) \sigma_x = \sum_{\{\boldsymbol{\sigma}\}} \prod_y \frac{1+m_y\sigma_y}{2} \sigma_x = \\
&\underset{(a)}{=} \sum_{\sigma_x = \pm 1} \Bigg(\prod_{y \neq x} \underbrace{\sum_{\sigma_y = \pm 1} \frac{1+m_y\sigma_y}{2}}_{1} \Bigg) \frac{1+m_x\sigma_x}{2} \sigma_x = \\
&= \sum_{\sigma_x = \pm 1} \sigma_x \frac{1+m_x\sigma_x}{2} = \frac{1+m_x}{2} - \frac{1-m_x}{2} = m_x
\end{aligned} \tag{1.9}$$

eqn:local-average

where in (a) we split the product in the case $y \neq x$ and $y = x$. Also note that the average is over ρ_0 and not the "true" pdf ρ.

Choice of parametrization. The distribution $\rho_x(\sigma_x)$ in (**??**) is the most general discrete distribution for a binary variable such as σ_x, just rewritten to highlight the average m_x.
In fact, consider a generic **binary** variable σ. Its distribution is:

$$\mathbb{P}[\sigma = +1] = p_+ \qquad \mathbb{P}[\sigma = -1] = p_-$$

Due to normalization, $p_+ + p_- = 1$, and so there is only **one free parameter** needed to completely specify the pdf:

$$\mathbb{P}[\sigma = +1] = p \qquad \mathbb{P}[\sigma = -1] = 1 - p$$

If we then rewrite p as function of the average $\langle \sigma \rangle = m$, we get:

$$m = \sum_{\sigma = \pm 1} \sigma \mathbb{P}[\sigma] = p - (1-p) = 2p + 1 \Rightarrow p = \frac{1+m}{2}$$

And so:

$$\mathbb{P}[\sigma = +1] = \frac{1+m}{2} \qquad \mathbb{P}[\sigma = -1] = \frac{1-m}{2}$$

Which can be rewritten more compactly as:

$$\rho(\sigma) = \frac{1+m\sigma}{2}$$

So we are not making any additional hypothesis other than that of a separable $\rho(\boldsymbol{\sigma})$ (given by the mean field approximation).

For simplicity, we work with βF_V, denoting $\beta J \equiv K$ and $\beta b_x \equiv h_x$. From the variational principle (**??**):

$$\beta F \leq \min_{\boldsymbol{m}} \beta F_V(\boldsymbol{m}, \boldsymbol{h}) = \min_{\boldsymbol{m}} \left(\beta \langle \mathcal{H} \rangle_0 - \frac{S[\rho_0]}{k_B} \right) \tag{1.10}$$

eqn:ising-variational

The average of $\mathcal{H}$ according to the ansatz is:

$$\langle \mathcal{H} \rangle_0 = \langle -J \sum_{\langle x,y \rangle} \sigma_x \sigma_y - \sum_x b_x \sigma_x \rangle_0 = -J \sum_{\langle x,y \rangle} \langle \sigma_x \sigma_y \rangle_0 - \sum_x b_x \langle \sigma_x \rangle_0$$

We already computed $\langle \sigma_x \rangle_0 = m_x$ in (**??**). For the two-point correlation, as ρ_0 is separable and thus σ_x and σ_y are decoupled, we get:

$$\langle \sigma_x \sigma_y \rangle_0 = \langle \sigma_x \rangle_0 \langle \sigma_y \rangle_0 = \sum_{\sigma_x} \frac{1+m_x\sigma_x}{2}\sigma_x \sum_{\sigma_y} \frac{1+m_y\sigma_y}{2}\sigma_y = m_x m_y$$

Thus:

$$\langle \mathcal{H}(\boldsymbol{\sigma}) \rangle_0 = -J \sum_{\langle x,y \rangle} m_x m_y - \sum_x b_x m_x = \mathcal{H}(\boldsymbol{m}) \qquad \texttt{eqn:H0avg} \tag{1.11}$$

This is valid more in general when applying the mean field approximation to even more complex Hamiltonians, as it is a consequence of the separability of ρ_0.

On the other hand, the entropy of ρ_0 can be directly computed. Noting that $\rho_x(\sigma_x)$ is exactly the same pdf we used in the non-interacting Ising Model, we can borrow the results (**??**) and (**??**, pag. **??**) from there:

$$\begin{aligned} -\frac{S[\rho_0]}{k_B} &= \sum_{\{\boldsymbol{\sigma}\}} \rho_0(\boldsymbol{\sigma}) \ln \rho_0(\boldsymbol{\sigma}) = \sum_x \sum_{\sigma_x} \frac{1+m_x\sigma_x}{2} \ln \frac{1+m_x\sigma_x}{2} = \\ &= \sum_x \left(\frac{1+m_x}{2} \ln \frac{1+m_x}{2} + \frac{1-m_x}{2} \ln \frac{1-m_x}{2} \right) \equiv \sum_x s_0(m_x) \qquad \texttt{eqn:rho0-ent} \end{aligned} \tag{1.12}$$

where we defined a *local entropy* s_0 as:

$$s_0(m) \equiv \frac{1+m}{2} \ln \frac{1+m}{2} + \frac{1-m}{2} \ln \frac{1-m}{2}$$

Substituting these results (**??**) and (**??**) back in (**??**) we arrive to:

$$\begin{aligned} \beta F_V(\boldsymbol{m}, \boldsymbol{h}) &= \beta H(\boldsymbol{m}) + \sum_x s_0(m_x) = \qquad \texttt{eqn:var-free-energy} \\ &= -K \sum_{\langle x,y \rangle} m_x m_y - \sum_x h_x m_x + \sum_x \left[\frac{1+m_x}{2} \ln \frac{1+m_x}{2} + \frac{1-m_x}{2} \ln \frac{1-m_x}{2} \right] \end{aligned} \tag{1.13}$$

where the first line holds for a generic Hamiltonian $\mathcal{H}(\boldsymbol{\sigma})$, and the second is specific for the Ising Model we are studying.

Then, we minimize $F_V(\boldsymbol{m}, \boldsymbol{h})$ with respect to $\boldsymbol{m}$, denoting the minimum as $F_V(\boldsymbol{M}, \boldsymbol{h})$:

$$\frac{\partial}{\partial m_x} \beta F_V \Big|_{\boldsymbol{m}=\boldsymbol{M}} \overset{!}{=} 0 \qquad \texttt{eqn:minimize} \tag{1.14}$$

$$\begin{aligned} 0 &\overset{!}{=} \frac{\partial}{\partial m_x} \left[-K \sum_{\langle x,y \rangle} m_x m_y - \sum_x h_x m_x + \sum_x \left(\frac{1+m_x}{2} \ln \frac{1+m_x}{2} + \frac{1-m_x}{2} \ln \frac{1-m_x}{2} \right) \right]_{\boldsymbol{m}=\boldsymbol{M}} = \\ &= -K \sum_{y \in \langle x,y \rangle} M_y - h_x + \frac{1}{2} \ln \frac{1+M_x}{2} + \frac{1+M_x}{\cancel{2}} \frac{2}{1+M_x} \frac{\cancel{1}}{2} - \frac{1}{2} \ln \frac{1-M_x}{2} - \frac{1-M_x}{\cancel{2}} \frac{2}{1-M_x} \frac{\cancel{1}}{2} = \\ &= -K \sum_{y \in \langle x,y \rangle} M_y - h_x + \frac{1}{2} \ln \left(\frac{1+M_x}{\cancel{2}} \frac{\cancel{2}}{1-M_x} \right) \end{aligned}$$

where the sum is over all nodes y neighbouring x, i.e. the ones included in some pair of neighbours $\langle y, x\rangle$ involving x.
Using the identity (**??**, pag. **??**)

$$\tanh^{-1} M_x = \frac{1}{2}\ln\frac{1+M_x}{1-M_x}$$

and rearranging leads to:

$$M_x(\boldsymbol{h}, K) = \tanh\left[K\sum_{y\in\langle y,x\rangle} M_y + h_x\right] \quad \text{eqn:variational-sol} \tag{1.15}$$

1.1.1 Physical meaning of the variational parameters M_x

It would be interesting to associate some physical meaning to the variational solution, and in particular understand what the M_x represent.

So, we found that:

$$\min_{\boldsymbol{m}} F_V(\boldsymbol{m}, \boldsymbol{h}) \equiv F_V(\boldsymbol{M}, \boldsymbol{h})$$

with the $\boldsymbol{M}$ given by solving the N equations (**??**), one for each node.

The *magnetization* given by the variational free energy is:

$$\begin{aligned}\langle\sigma_x\rangle_V &\underset{(??)}{=} -\frac{\partial}{\partial h_x}[\beta F_V(\boldsymbol{M}, \boldsymbol{h})] = -\beta\left[\underbrace{\sum_y \frac{\partial F_V}{\partial m_y}(\boldsymbol{m}, \boldsymbol{h})\frac{\partial m_y}{\partial h_x}}_{0\,(??)} - \underbrace{\frac{\partial F_V}{\partial h_x}(\boldsymbol{m}, \boldsymbol{h})}_{M_x\,(??)}\right]_{\boldsymbol{m}=\boldsymbol{M}} = \\ &= M_x \end{aligned} \quad \text{eqn:MX-meaning} \tag{1.16}$$

Note that the variational free energy F_V **is not** the *ansatz free energy* F_0, and so $\langle\sigma_x\rangle_V$ and $\langle\sigma_x\rangle_0$ are different averages, and (**??**) should not be confused with (**??**).

So, M_X is the best estimate of the *true magnetization* σ_x, as it is obtained with the F_V *closest* to the real F.

1.1.2 Uniform case

Suppose the magnetic field is uniform $h_x \equiv h$. In this case, the system is **translationally invariant**. So, it is reasonable to consider the *ansatz* where also all the local magnetizations are the same: $m_x \equiv m$, and search for a single value of m.

Given these assumptions, (**??**) becomes:

$$\beta F_V(m, h) = -Km^2\sum_{\langle x,y\rangle} 1 - mh\sum_x 1 + \left[\frac{1+m}{2}\ln\frac{1+m}{2} + \frac{1-m}{2}\ln\frac{1-m}{2}\right]\sum_x 1$$

Then $\sum_x 1$ is just the number of nodes N, and $\sum_{\langle x,y\rangle} 1$ is the number of possible pairs, which is Nd for a d-dimensional cubic lattice (each node contributes with one pair for every possible *direction*). Dividing by N:

$$\beta\frac{F_V(m, K, h)}{N} = -Kdm^2 + \frac{1+m}{2}\ln\frac{1+m}{2} + \frac{1-m}{2}\ln\frac{1-m}{2} - hm \quad \text{eqn:FV-uniform} \tag{1.17}$$

The equation for M_X (**??**) becomes:

$$M(h,K) = \tanh\left[KM\sum_{y\in\langle y,x\rangle} 1 + h\right]$$

The sum is over all *neighbours* of x, which are $2d$ for a d-dimensional cubic lattice (2 for every *direction*), leading to:

$$M(h,K) = \tanh(2dKM + h) \tag{1.18}$$

eqn:uniform-variational-eq

A. No external field

Case 1. $h = 0$

Let's start with the case of no external field $h = 0$. In this case, the variational free energy (**??**) is an **even** function of m:

$$F_V(m,0) = F_V(-m,0)$$

We can then study the solutions of (**??**):

$$M = \tanh(2dKM) \qquad M(K,0) \equiv M(K) \tag{1.19}$$

eqn:h0case

Clearly $M = 0$ is always a solution. Depending on $2dK$, there can be two more solutions, as can be seen by plotting each side and looking for intersections (**??**).

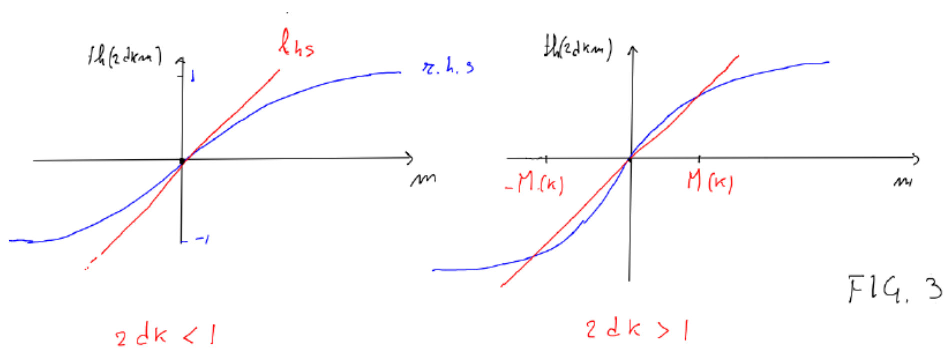

Figure (1.1) – Solutions of (**??**) are intersections of the two curves. fig:uniformh0

The plots in (**??**) can be obtained by expanding $\tanh x$ in Taylor series around $x = 0$. The first three derivatives are:

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}x}\tanh x &= 1 - \tanh^2 x\\
\frac{\mathrm{d}^2}{\mathrm{d}x^2}\tanh x &= -2\tanh x(1-\tanh^2 x)\\
\frac{\mathrm{d}^3}{\mathrm{d}x^3}\tanh x &= -2(1-\tanh^2 x) + 4\tanh^2 x(1-\tanh^2 x)
\end{aligned}$$

So:

$$\tanh x = \tanh 0 + x\frac{\mathrm{d}}{\mathrm{d}x}\tanh x\Big|_{x=0} + \frac{x^2}{2}\frac{\mathrm{d}^2}{\mathrm{d}x^2}\tanh x\Big|_{x=0} + \frac{x^3}{3!}\frac{\mathrm{d}^3}{\mathrm{d}x^3}\tanh x\Big|_{x=0} + \cdots =$$

$$= x - \frac{2x^3}{3 \cdot 2 \cdot 1} + O(x^5) = x - \frac{x^3}{3} + O(x^5) \tag{1.20}$$

eqn:tanh-exp

For small x, $\tanh x$ is linear, and in particular $\tanh(2dKM)$ is a line passing through the origin with slope $2dK$. It that slope is **less** than te one of $y = M$, i.e. 1, then the only intersection is at $M = 0$ (left of fig. **??**). However, if $2dK > 1$, then there will be two other solutions (right of fig. **??**).
In summary:

- $2dK < 1 \Rightarrow K < K_c \equiv 1/2d$, (**??**) has only one solution $M = 0$.
- If $2dK > 1 \Rightarrow K > K_c$, there are 3 solutions: $M = 0, \pm M(K)$.

In the case $K > K_c$, we need to understand which of the three solution leads to the absolute minimum of F_V. So, let's proceed by expanding $\beta F_V(m, 0)/N \equiv f(m)$ (**??**) for small m. The first four coefficients are:

$$\begin{aligned}
f(0) &= \frac{1}{2}\ln\frac{1}{2} + \frac{1}{2}\ln\frac{1}{2} = -\frac{1}{2}\ln 2 - \frac{1}{2}\ln 2 = -\ln 2 \\
f'(0) &= -2Kdm + \cancel{\frac{1}{2}\ln\frac{1+m}{2}} + \cancel{\frac{1}{2}} - \cancel{\frac{1}{2}\ln\frac{1-m}{2}} - \cancel{\frac{1}{2}}\Big|_{m=0} = 0 \\
f''(0) &= -2Kd + \frac{1}{4}\frac{2}{1+m} + \frac{1}{4}\frac{2}{1-m}\Big|_{m=0} = 1 - 2Kd \\
f^{(3)}(0) &= -\frac{1}{2(1+m)^2} + \frac{1}{2(1-m)^2}\Big|_{m=0} = 0 \\
f^{(4)}(0) &= -\frac{1}{2}\frac{-2}{(1+m)^3} + \frac{1}{2}(-2)\frac{-1}{(1-m)^3}\Big|_{m=0} = 2
\end{aligned}$$

Clearly all odd terms vanish because $F_V(m, 0)$ is **even**. Then:

$$\begin{aligned}
\frac{\beta F_V(m, h=0)}{N} &= f(0) + m f'(0) + \frac{m^2}{2} f''(0) + \frac{m^3}{3!} f^{(3)}(0) + \frac{m^4}{4!} f^{(4)}(0) + \dots = \\
&= -\ln 2 + \frac{1 - 2Kd}{2} m^2 + \frac{m^4}{12} + O(m^6)
\end{aligned}$$

Let's focus on the highlighted quadratic term. We distinguish three cases:

1. When $2Kd < 1$ ($K < K_c$) the coefficient its positive, meaning that, for $x \sim 0$, F_V behaves like a convex parabola (left of fig. **??**). As $K = \beta J = J/k_B T$, this holds for $T > T_c = 2dJ/k_B$, where T_c is called the system's **critical temperature**.

   Note how, in this case, the variational free energy has a single global minimum at $m = 0$.

2. Now, if we let $2Kd = 1$ ($K = K_c = 1/2d$, or $T = T_c = 2dJ/k_B$), then the quadratic coefficient vanishes, and for $m \sim 0$ the variational free energy has the shape of a *quartic* (m^4), meaning that it is close to 0 and "very flat" for $m \to 0$. Still, there is only one global minimum at $m = 0$.

3. However, if $2Kd > 1$, then F_V is like a **concave** parabola near the origin. So $m = 0$ becomes a local maximum, and $m = \pm M(K)$ are two equivalent local minima.

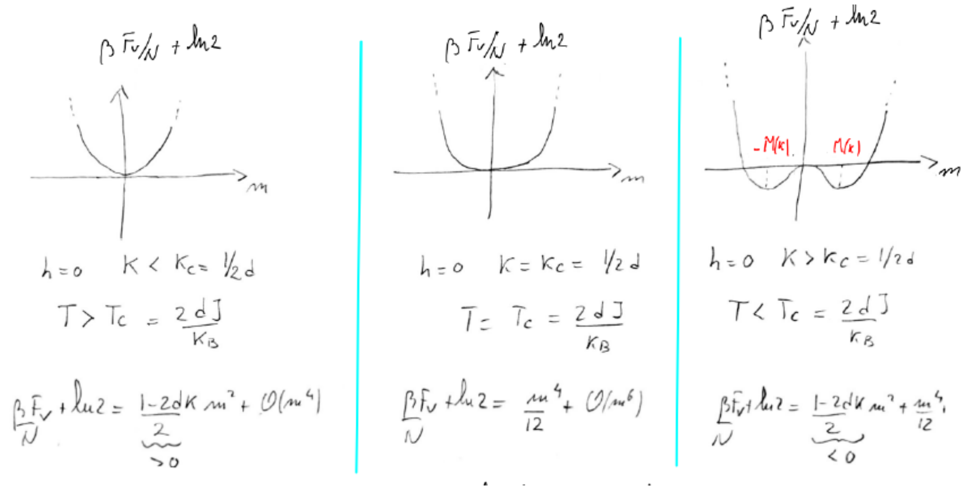

Figure (1.2) fig:variational_cases

Thus, depending on the **temperature**, the system's behaviour changes *fundamentally.*

Once we have found the solution M for the minimum, the **best estimate** of the exact *free energy* βF is given by **??** evaluated at $m = M$ and $h = 0$:

$$\beta \frac{F_V(M,H,0)}{N} = -KdM + \frac{1+M}{2}\ln\frac{1+M}{2} + \frac{1-M}{2}\ln\frac{1-M}{2}$$

Physical meaning of $M(K)$

$M(K)$ and the spontaneous magnetization

When $T < T_c$, we found that the free energy is best approximated by a function with two local minima at $\pm M(K)$ - which we have interpreted as estimates of the system's **magnetization**. So, this mechanism could explain the experimentally observed phenomenon of **spontaneous magnetization**.

Explicitly, we defined the spontaneous magnetization *per node* m_S (**??**) as:

$$-\lim_{h\downarrow 0}\frac{1}{N}\lim_{N\uparrow\infty}\frac{\partial}{\partial h}(\beta F) = \lim_{h\downarrow 0}\langle\frac{\sum_x \sigma_x}{N}\rangle = m_S \qquad \text{eqn:ms} \quad (1.21)$$

In particular, the thermodynamic limit must be taken **before** the $h \to 0$ limit. We can now use the variational free energy to compute an estimate of m_S. Note that in (**??**), the free energy density *does not* depend on N, so the limit of $N \to \infty$ is trivial. Then we just need to differentiate with respect to h and set $m = M$, the minimum found by solving (**??**). Thus, the *variational estimate* of m_S is given by:

$$\begin{aligned} m_S\Big|_{\text{var.}} &= -\lim_{h\downarrow 0}\frac{\partial}{\partial h}\frac{F_V(M,K,h)}{N} = -\lim_{h\downarrow 0}\left[\underbrace{\frac{\partial F_V}{\partial m}(m,K,h)}_{0\ (??)}\frac{\partial M}{\partial h} + \underbrace{\frac{\partial F_V}{\partial h}(m,K,h)}_{-m\ (??)}\right]_{m=M} = \\ &= \lim_{h\downarrow 0} M(K,h) = M(k) \qquad \text{eqn:variational-spontaneous-magnetization} \quad (1.22) \end{aligned}$$

where $M(K,h)$ is the solution of (**??**), which, in the limit $h \to 0$, becomes one of the solutions we found in the $h = 0$ case, since it is an analytic function. So $m_S = 0$ if $2dK < 1$, and $\neq 0$ otherwise.

We can then study how the solution $M(K)$ of (**??**) varies as a function of $K^{-1} = k_B T/J$. This can be done numerically - but to get some understanding we consider the case near criticality $K \approx K_c = 1/2d$. From fig. **??** and fig. **??** we expect $M \approx 0$ when $K \approx K_c$.

So, using the expansion of $\tanh x$ (**??**) for small x, (**??**) becomes:

$$M = 2dKM - \frac{(2dK)^3 M^3}{3} + O(M^5)$$

One solution is clearly $M = 0$, $\forall K$.

$M(K)$ near criticality

For the other **solutions**, we suppose that $K > K_c = 1/2d$, e.g. $K = K_c + \delta$ with $\delta \approx 0^+$, and then divide by M to get:

$$\begin{aligned} M^2 &= \frac{3}{(2dK)^3}(2dK - 1) + O(M^4) = \\ &= \frac{6d}{(K/K_c)^3}(K - K_c) + O(M^4) = \\ &= \frac{6d}{[(K_c + \delta)/K_c]^3}(\cancel{K_c} + \delta - \cancel{K_c}) + O(M^4) = \\ &= 6d \frac{\delta}{(1 + \delta/K_c)^3} + O(M^4) = \\ &= 6d\delta + O(\delta^2) \end{aligned}$$

For $\delta \approx 0$, $\delta/(1 + \delta/K_c)^3 \approx \delta$, and so M^2 is of order δ, meaning that M^4 is of order δ^2.

Taking the square root:

$$M(K) = \sqrt{6d}(K - K_c)^\beta + O(K - K_c) \quad \texttt{eqn:mean-field-MS} \tag{1.23}$$

Critical exponent and universality

where $\beta = 1/2$ is the **critical exponent**. Note that the behaviour of the spontaneous magnetization near criticality is given by a power law in the distance to the critical point K_c: this happens more in general, not only for the Ising Model, and does not depend on the details of the model (**universality**).

(**??**) also produces a **singularity** at $K = K_c$, where $M(K)$ starts rising from 0 in a non-smooth manner (fig. **??**).

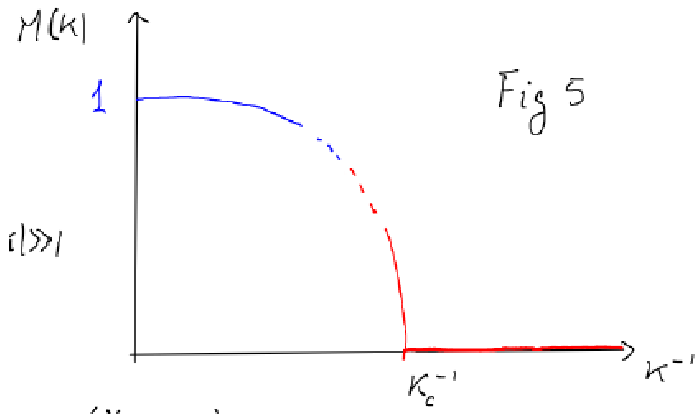

Figure (1.3) – Plot of the spontaneous magnetization $M(K)$ (estimated from the variational free energy) as function of temperature ($K^{-1} \propto T$). From fig. **??** we know that $M(K) = 0$ for $K < K_c$. The red curve at $K \approx K_c$ is given by (**??**), while the blue curve at $K \to \infty$ derives from (**??**) Note the **singularity** at $K = K_c$, the critical point. `fig:MK_plot`

The result in (**??**) is an estimate given by the mean field approximation. However, the same kind of relation *holds* in the true model, just with a different exponent β. For the $d = 2$ case, $\beta = 1/8$ can be exactly determined, while for $d > 2$ one resorts to numerical methods, obtaining $\beta \approx 0.31$ at $d = 3$, and - surprisingly - $\beta = 1/2$ for $d > 3$. Again, this is not a specific behaviour: the mean field approximation happens to become **exact** in $d \geq 4$ in many cases, as we will see later on.

The validity of the mean field approximation

If we instead study the behaviour at low temperatures ($K \gg 1$), we expect from fig. **??** to see $M \approx 1$, meaning that the argument $2dKM(k)$ of the tangent in (**??**) becomes very large. So we expand $\tanh x$ accordingly:

$$\tanh x = \frac{e^x - e^{-x}}{e^x + e^{-x}} \frac{e^{-x}}{e^{-x}} = \frac{1 - e^{-2x}}{1 + e^{-2x}} \underset{(a)}{=} (1 - e^{-2x})(1 - e^{-2x} + e^{-4x} + \dots) =$$
$$= 1 - 2e^{-2x} + 2e^{-4x} + O(e^{-6x})$$

where in (a) we used the geometric series expansion:

$$\frac{1}{1+x} = 1 - x + x^2 - x^3 + \dots$$

And substituting in (**??**) we get:

$$M(K) = 1 - 2e^{-4dKM(k)} + O(e^{-8KdM(k)}) \underset{(b)}{=} 1 - 2e^{-4dK} + O(e^{-12dK}) \quad \text{eqn:low-temeperature-var} \tag{1.24}$$

where in (b) we substituted $M(k) \approx 1$ in the right side, noticing that all other terms are of order e^{-12dK} or higher. This result agrees with the low temperature expansion we did in the $d = 2$ case in (**??**, pag. **??**). So the spontaneous magnetization quickly approaches 1 when $K^{-1} \to 0$ ($T \to 0$).

B. External field

If $h \neq 0$, from (**??**) we have:

2. Case $h \neq 0$

$$\beta \frac{F_V(m, K, h)}{N} = \beta \frac{F_V(m, K, 0)}{N} - hm$$

So the variational equations (**??**) become:

$$h = \frac{\partial}{\partial m} \left[\beta \frac{F_V(m, K, 0)}{N} \right]_{m=M} = (\tanh^{-1} m - 2dKm) \Big|_{m=M} \quad \text{eqn:var-eq-h} = \tag{1.25}$$
$$\underset{M \approx 0}{=} M(1 - 2dK) + \frac{M^3}{3} + \frac{M^5}{5} + \frac{M^7}{7} + \dots$$

Depending on the sign of $1 - 2dK$, i.e. if $2dK$ is lower of higher than 1, the slope at the origin will be either positive or negative, leading to the plots in fig. **??**.

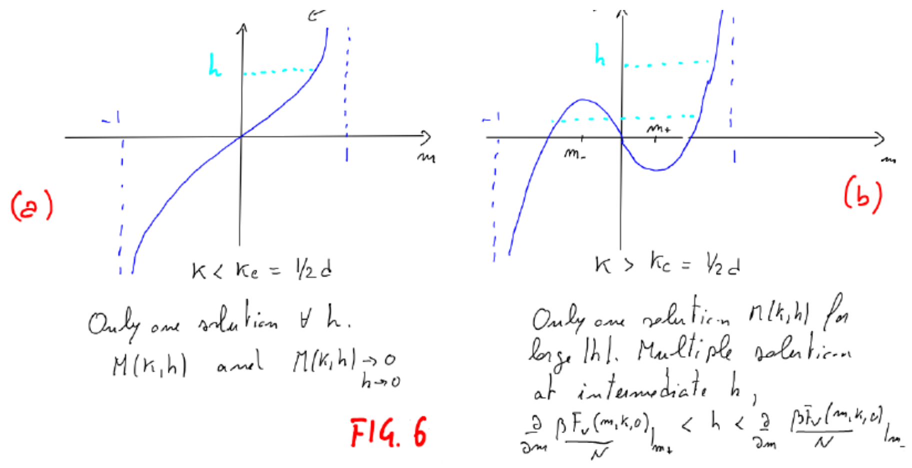

Figure (1.4) – Plot of the right hand side of (**??**), i.e. the variational estimate of magnetization, as function of m. `fig:hplot`

So there are two cases:

1. If $K < K_c = 1/2d$, then the right side of (**??**) is strictly increasing, and so admits only one intersection with an horizontal line $y = h$, meaning that there is only one solution for $M(h, K)$ (in general $\neq 0$). If we then let $h \to 0$, $M(K, h) \to 0$ smoothly, and so $m_S = 0$, as expected.

2. If $K > K_c$, instead, the plot is the one on the right of fig. **??**, and multiple intersections with $y = h$ are possible if h lies in a certain range:
$$\frac{\partial}{\partial m} \frac{\beta F_V(m, K, 0)}{N}\Big|_{m_+} < h < \frac{\partial}{\partial m} \frac{\beta F_V(m, K, 0)}{N}\Big|_{m_-}$$
where $m_\pm$ are the local minima/maxima of the right side of (**??**).

In the $K > K_c$ case, in order to understand which of the possible multiple solutions $\{M_i\}_{i=1,2,3}$ corresponds to the minimum of F_V we refer to fig. **??**.

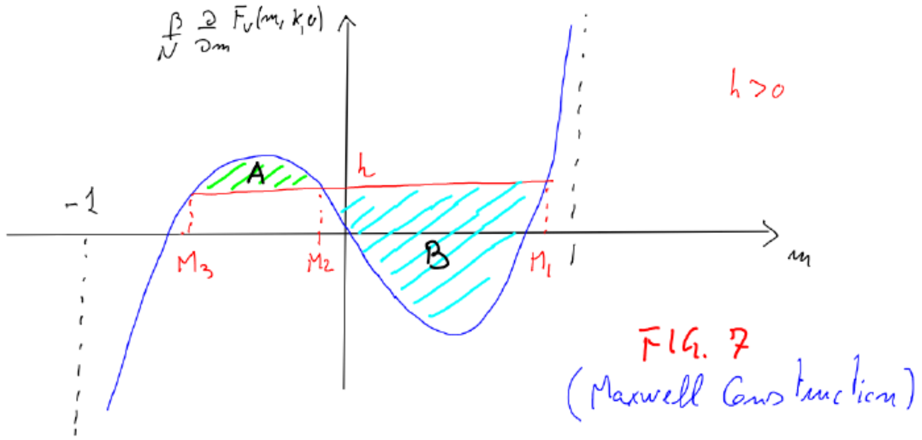

Figure (1.5) `fig:variational_energy_h`

To simplify notation, let's denote as f_i the free variational energy evaluated at a solution M_i:
$$f_i = \frac{\beta F_V(M_i, K, h)}{N} = \frac{\beta F_V(M_i, K, 0)}{N} - hM_i$$

Then note that differences of f_i can be rewritten as integrals, which can be roughly evaluated by looking at fig. **??**. Then, for $h > 0$:

$$f_1 - f_2 = \int_{M_2}^{M_1} \left(\frac{\beta}{N} \frac{\partial}{\partial m} F_V(m, K, 0) - h \right) \mathrm{d}m = -\text{Area of } \mathbf{B} < 0 \Rightarrow f_1 < f_2$$
$$f_2 - f_3 = \int_{M_3}^{M_2} \left(\frac{\beta}{N} \frac{\partial}{\partial m} F_V(m, K, 0) - h \right) \mathrm{d}m = -\text{Area of } \mathbf{A} < 0 \Rightarrow f_2 < f_3$$
$$f_1 - f_3 = \int_{M_3}^{M_1} \left(\frac{\beta}{N} \frac{\partial}{\partial m} F_V(m, K, 0) - h \right) \mathrm{d}m = \text{Area of } \mathbf{A} - \text{Area of } \mathbf{B} < 0 \Rightarrow f_1 < f_3$$

Summarizing:

1. For $h > 0$, the area of **B** is always bigger than that of **A**. So, at the end, $f_1 < f_2 < f_3$.
2. For $h = 0$, the two areas **A** and **B** become equal, and f_1 and f_3 are two degenerate minima.
3. On the other hand, if $h < 0$, all inequalities are reversed, and $f_3 < f_2 < f_1$. So, when h changes sign, the system *jumps* to a different minimum.

Intuitively, a $h > 0$ leads to a *preference* for a positive magnetization, and, conversely, $h < 0$ for a negative magnetization.

A plot of the solution $M(K, h)$ corresponding to the minimum of F_V as a function of h is shown in fig. **??**.

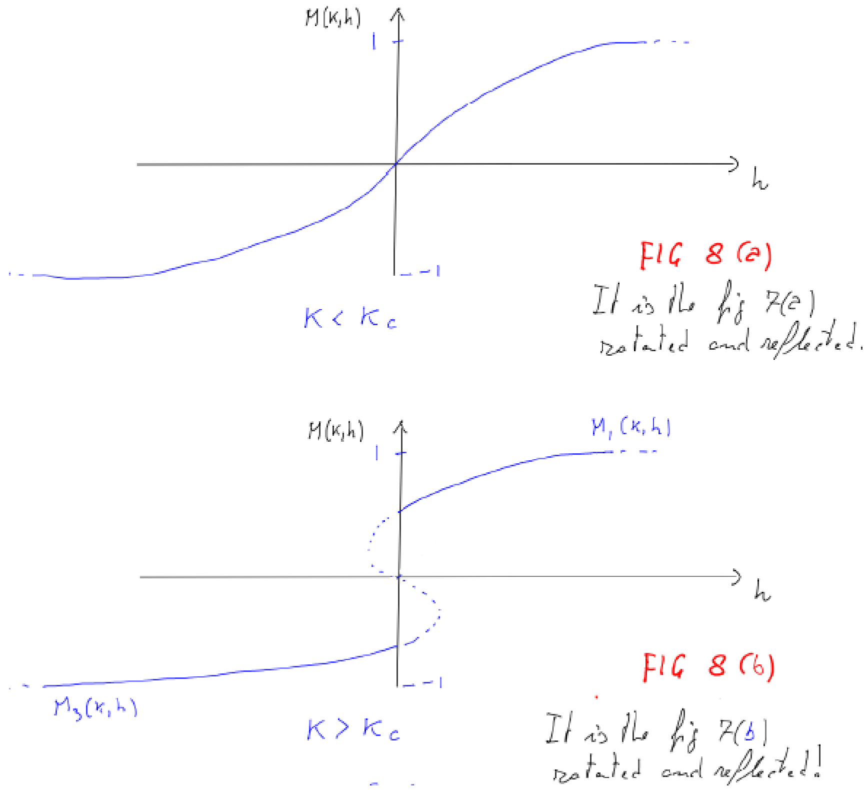

Figure (1.6) – Plot of $M(K,h)$ (variational estimate of magnetization, obtained by minimizing F_V) as a function of the external field h, which can be obtained by rotating and reflecting fig. **??**. If $K < K_c$ (top) the magnetization varies continuously as a function of h. If $K > K_c$, instead, (bottom) there is a discontinuity at $h = 0$, given by the system's transition to a different minimum (M_3 instead of M_1) `fig:Mhplot`

All of these results about criticality are summarized in fig. **??**.

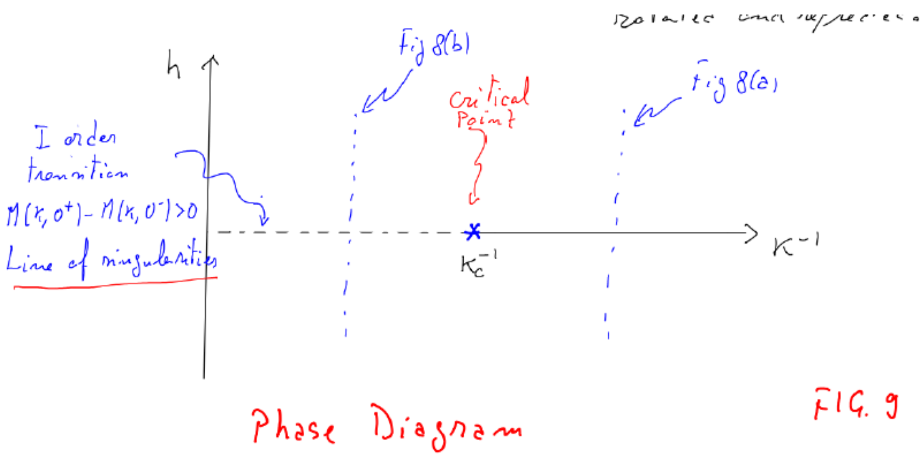

Figure (1.7) – Phase diagram representing all the singular points of $M(K,h)$ as a dashed line. Any curve surpassing the dashed part (left of K_c^{-1}) has a discontinuity (first-order transition). One such path is the one in the bottom plot of fig. **??**. On the other hand, a curve surpassing $h=0$ *at the right* of K_c^{-1}, however, is smooth; and one such example is given by the top curve of fig. **??**. So, starting at a point (h, K^{-1}) with $h>0$, we can construct *two kinds* of paths arriving to the phase with $h<0$: one passing through a high-temperature state and without phase-transitions, and one with a phase-transition at a low temperature. Something analogous happens for the vapour-liquid transition: it can be observed as an abrupt change (phase transition) at sufficiently low temperatures, or as a completely smooth process if pressure is increased such that phase differences are removed (the “gas looks like a liquid”). `fig:phase-diagram-uniform`

We conclude by stressing that the **singularities** at $h=0$ and $K>K_c$ *emerge* from to the variational principle as a consequence of the minimization.

Remarks on the mean-field approximation. The Mean Field (MF) model predicts a phase transition in all $d>0$. However we know that this is not true in $d=1$, where no phase transition is observed (pag. **??**). Still, for $d>1$ the MF is at least qualitatively correct. Impressively, such a simple model agrees *exactly* with simulation at $d \geq 4$, at least for the behaviour of magnetization near criticality.

Mean Field and symmetry breaking. For $h=0$, the Ising Model Hamiltonian:

$$\mathcal{H}(\boldsymbol{\sigma}) = -J \sum_{\langle x,y \rangle} \sigma_x \sigma_y$$

is **symmetric** with respect to the transformation $\sigma_x \to -\sigma_x \ \forall x$, i.e. $\mathcal{H}(\boldsymbol{\sigma}) = -\mathcal{H}(\boldsymbol{\sigma})$. In any **finite** system ($N<\infty$), this symmetry implies that $\langle \sigma_x \rangle = -\langle \sigma_x \rangle \Rightarrow \langle \sigma_x \rangle = 0$, meaning that no spontaneous magnetization can be observed. However, in the **infinite volume**, this symmetry is **spontaneously broken** below some critical temperature and $\langle \sigma_x \rangle \neq 0$.
We have shown how this occurs in the mean field approximation. Specifically, the symmetry that is broken for the Ising model is $\mathbb{Z}_2$.

If we instead consider the Hamiltonian:

$$H(\boldsymbol{\sigma}) = -J \sum_{\langle x,y \rangle} \boldsymbol{\sigma_x} \cdot \boldsymbol{\sigma_y}$$

where $\boldsymbol{\sigma_x} \in \mathbb{R}^n$ and $\|\boldsymbol{\sigma_x}\| = 1$, then the group symmetry is $\mathrm{O}(n)$, the orthogonal group, and $H(R\boldsymbol{\sigma}) = H(\boldsymbol{\sigma})$, where R is a $n \times n$ matrix such that $\|R\boldsymbol{\sigma}\| = \|\boldsymbol{\sigma}\|^2 = 1$, i.e. a orthogonal ("rotation") matrix satisfying $R^T R = RR^T = \mathbb{I}$. There are rigorous results establishing that discrete symmetries like $\mathbb{Z}_2$ cannot be spontaneously broken in $d = 1$ (Landau arguments) whereas continuous symmetries, like $\mathrm{O}(n)$, cannot be spontaneously broken in $d \leq 2$ (Mermin-Wagner theorem). In both cases only short-range interactions are assumed.