Fermionic renormalization group flows – Technique and Theory

Manfred Salmhofer[a], Carsten Honerkamp[b]

[a] Mathematik, ETH–Zentrum, 8092 Zürich, Switzerland

[b] Theoretische Physik, ETH–Hönggerberg, 8093 Zürich, Switzerland

January 23, 2003

Abstract

We give a selfcontained derivation of the differential equations for Wilson's renormalization group for the one–particle–irreducible Green functions in fermionic systems. The application of this equation to the (t, t')– Hubbard model has appeared in [9]. Here we focus on theoretical aspects. After deriving the equations, we discuss the restrictions posed by symmetries on the effective action. We discuss scaling properties due to the geometry of the Fermi surface, and give precise criteria when they justify the use of one–loop flows. We also discuss the relationship of this approach to other RG treatments, as well as aspects of the practical treatment of truncated equations, such as the projection to the Fermi surface and the calculation of susceptibilities.

1 Introduction

Renormalization group (RG) studies of fermionic models are important for understanding models of solid-state theory and high–energy physics. In this paper, we set up a system of equations suitable for studying flows for general fermionic systems, in particular ones with a Fermi surface at any density and discuss general, but nontrivial, properties of their solutions. The system of equations applies both to "normal" and to symmetry–broken situations of fermionic systems with short–range interactions in $d \geq 1$ (by short–range we mean that the two–particle interaction decays so fast that its Fourier transform is a regular function). In [9] we have, together with N. Furukawa and T.M. Rice, applied the RG equations that we derive here to the two–dimensional (t, t')–Hubbard model in the regime relevant for high–temperature superconductivity. The purpose of the present paper is to give more background on the theoretical aspects of the equations, in particular on a detailed argument which justifies the one–loop flow in a certain scale range even if the scale–dependent interaction is not small any more. In the following, we give an overview; the details are filled in in the later sections.

1.1 The RG and the three scale regimes

The Wilsonian RG organizes the functional integration corresponding to the grand canonical trace as an iterated integral over degrees of freedom with energies in a certain range, i.e. those that belong to a corresponding length scale. As such, it is simply a rewriting of the generating functional of the correlation functions as a function of an energy scale ϵ, which, in contrast to other RG schemes, does not require any assumptions about perfect scaling laws as a function of ϵ. This idea of integrating over some degrees of freedom does a priori not require a splitting of the Hamiltonian into a term describing free particles and an interaction term, with an assumption that the interaction term is small. However, it is of course in general a very difficult task to calculate renormalization group flows in strongly coupled systems, and we shall assume that the interaction in the Hamiltonian, which plays the role of the initial interaction, is weak, so that we can think of energy scales as *kinetic* energy scales. Thus the method we describe here is a weak–coupling technique. A well–known property of the effective interaction is that it cannot be described by a single coupling constant, and in the cases

we are interested in, we need to study a coupling function that depends on the momenta of the particles, as well as on the energy scale ϵ. As the energy scale ϵ decreases, the generic behaviour in two or more spatial dimensions is that the coupling function grows. Thus, even with a weak initial interaction, the flow eventually leaves the weak–coupling regime, unless the temperature is above all critical temperatures, so that it prevents any instabilities, see [2] and the discussion below. This growth of the effective interaction signals instabilities, and from the technical point of view it means that one has to switch to a different description of the system, e.g. in terms of composite fields describing Cooper pairs or spin variables. Because of the semigroup property of the RG flow (which we recall in Section 2.2), one can stop the flow at a nonzero scale ϵ_2 and rewrite the effective interaction in terms of the new fields at that scale. However, in the context of the present method, the first question is to what scale one can get using the weak–coupling flow. In Section 5 we give a precise criterion of how large the coupling constants are allowed to get before the corrections to the one–loop flow become significant. It turns out that there are nontrivial effects that suppress these corrections even if the couplings are not small any more, and these effects depend on the geometry of the Fermi surface. In the following, we describe briefly the main idea behind a picture of three scale regions.

The renormalization group defines an effective interaction for the fermions with kinetic energy $e(\mathbf{p})$ less than some ϵ, obtained from the original interaction of the model by integrating over the degrees of freedom with energies bigger than ϵ (in the technical implementation, it is useful to use smooth cutoff functions). The phase space $\{\mathbf{p} : |e(\mathbf{p})| \leq \epsilon\}$ of the fermions with energy at most ϵ is a neighbourhood of the Fermi surface. Thus the shape of the Fermi surface determines the low–energy phase space, and this essentially determines which terms are dominant in the RG equations. If the Fermi surface is curved, there are power counting gains that suppress contributions from all graphs that have overlapping loops. This is made precise by an inequality for two– and higher–loop integrals that gives a rigorous bound for such integrals [11, 12, 14]. The constants in that bound depend on the curvature of the Fermi surface, and they diverge when the curvature of the Fermi surface vanishes. As shown in [11, 2], the only four–legged graphs that do not have overlapping loops are those that are generated by the one–loop RG equation. This implies that there is an energy scale ϵ_1, *determined by the curvature of the Fermi surface*, below which "overlapping loop" effects [11]

produce small scale factors that suppress corrections to the one–loop flow. Thus, below this energy scale, one has an argument for the validity of the one–loop flow, which does not require coupling constants to be very small (see Section 5).

The flow thus splits in a natural way in three energy scale regions, as follows. Regime 1 are the high scales, with energy above ϵ_1. In this regime, the only justification of a weak–coupling treatment is that the coupling constants are indeed small, so that higher order terms are small. On the other hand, this justification is mathematically rigorous: because the energy scale ϵ_1 provides an infrared cutoff for the functional integral for the effective action at scale ϵ_1, and because the fields are fermionic, there are rigorous proofs [13] that perturbation theory for the effective action at scale ϵ_1 converges. Thus it can be used to calculate the effective action at scale ϵ_1 reliably, if the coupling is smaller than a value that depends on ϵ_1.

Regime 2 is below ϵ_1. The scaling improvements due to overlapping loops now supress corrections to the one–loop flow. The latter usually gives growing coupling constants, especially in the physically interesting cases, which signal the tendency towards (possibly competing) instabilites. However, the scaling improvements due to overlapping loops are strong enough to suppress the corrections to the one–loop flow even when the coupling constants are not small any more. The one–loop flow itself leads to a divergence of the coupling constants at a certain scale ϵ^*. Thus, at a scale $\epsilon_2 > \epsilon^*$, the coupling constants become so large that the scaling improvements are insufficient to suppress the large factor from the coupling constants, and then the one–loop approximation breaks down. The scales ϵ_1 and ϵ_2 will be determined in detail in Section 5; see in particular eq. (111).

Regime 3 is below ϵ_2. Here, the theory is strongly coupled. However, the interaction has changed - and usually simplified - during the flow. For instance, in the superconducting case, a leading term has emerged which corresponds to the BCS model. In the (t, t') Hubbard model, the interaction has also developed a pronounced k–space structure at that scale [9], which may lead to a tractable theory, but a full strong–coupling theory has not yet been developed.

In summary, if one starts out at weak coupling, the one–loop RG flow is justified down to an energy scale ϵ_2 which is determined by the curvature of the Fermi surface, and at which the effective coupling function need not

be small any more. The details of the flow down to scale ϵ_2 depend on the shape of the Fermi surface, as well as on the strength of the initial coupling constant. We give a brief discussion of some cases in the next subsection.

1.2 Some consequences

1.2.1 The Kohn–Luttinger effect

In the case where the Fermi surface is positively curved and where the Umklapp scattering is irrelevant, the three–scale picture provides a simple view of the Kohn–Luttinger effect at weak coupling. Because the Fermi surface is curved, the particle–hole terms in the flow equation (67) are small compared to the particle–particle terms. The leading part of the coupling function is then obtained by projecting on the Fermi surface, and one can expand the coupling function in an orthonormal basis of functions ϕ_l on the Fermi surface. The particle–particle flow has the property that it decouples different l, so the corresponding coupling constants g_l all flow independently, according to the equations

$$\dot{g}_l = -\tilde{\beta}_l g_l^2 \tag{1}$$

(for a more detailed discussion, see [3], Section 4.5). Because (1) is only accurate below ϵ_1, the initial condition for the g_l is given at the scale ϵ_1. If g_l is a repulsive coupling at ϵ_1, i.e. $g_l(\epsilon_1) > 0$, then g_l decreases as ϵ decreases. If g_l is an attractive coupling at ϵ_1, i.e. $g_l(\epsilon_1) < 0$, then g_l increases in absolute value (see Figure 1). The Kohn–Luttinger effect is the observation that for a weak repulsive interaction at scale ϵ_0, there are l for which g_l is negative at the lower scale ϵ_1, and thus gets larger in absolute value below that scale (see Figure 1), eventually leading to a superconducting state with the gap symmetry given by l.

Because the maximal four–point function defined in Section 5 initially grows at most logarithmically in the energy scale, renormalized perturbation theory above a scale ϵ converges if the biggest initial coupling constant g_0 at scale ϵ_0 satisfies an inequality $|g_0| \log \epsilon_0/\epsilon \ll 1$ (here "renormalized" means that the Fermi surface shift is treated appropriately [11, 12]).

Thus, if $|g_0| \log \epsilon_0/\epsilon_1 \ll 1$, one can determine the superconducting instabilities of the weak–coupling model by doing a perturbative calculation to determine the g_l at scale ϵ_1 and checking which of these coupling constants

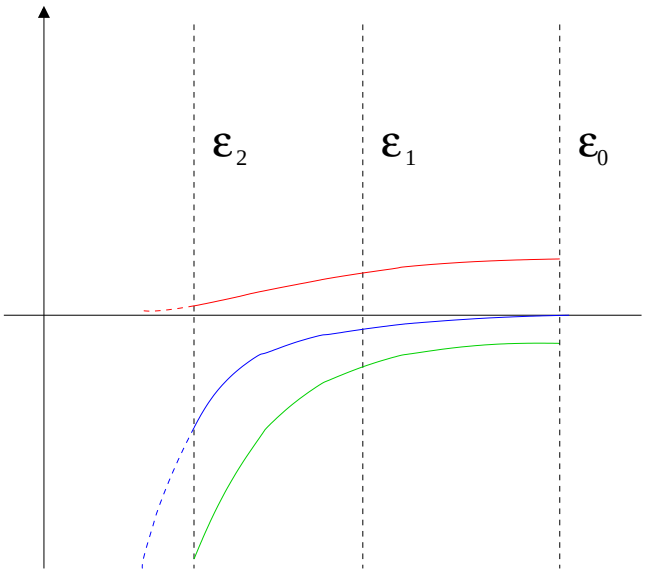


Figure 1: The three scales and the flow of some coupling constants. The flow starts at scale ϵ_0. Below scale ϵ_1, but above ϵ_2, scaling effects from overlapping loops justify a one–loop flow even if the couplings are not small. The flow of the coupling constants is sketched for the situation without Umklapp scattering and where the particle–particle term dominates the flow. In that case, couplings that are negative grow in absolute value, positive couplings get suppressed. Below scale ϵ_2, the one–loop flow cannot be justified any more.

is the most negative.

It is important to note that the above–mentioned decoupling of the flow of the coupling constants g_l holds only in absence of Umklapp scattering and when the particle–hole diagrams are not singular (this requires in particular the absence of van Hove singularities). If the particle–hole terms in the flow are important, the above–mentioned expansion in Fermi surface harmonics ϕ_l does not lead to decoupled flow equations. This, and the Umklapp scattering, is physically very important in the (t, t')–Hubbard model (see [16, 9], and the references therein).

1.2.2 Small curvature and nesting effects

The Kohn–Luttinger effect always drives the system to a superconducting state if the coupling function stays small down to the scale where the particle–particle flow is leading. As discussed above, such a scale always exists if the Fermi surface is curved, but it goes to zero when the curvature goes to zero. By the above reasoning, if the curvature of the Fermi surface is nonzero, one can make the initial coupling constant so small that one stays in the weak–coupling regime down to the scale where the curvature effects become relevant. Thus, if the Fermi surface is regular and curved, and if coupling constant is chosen small enough, *depending on the Fermi surface and thus in particular on the electron density*, the instability will always be of a superconducting type.

However, if the initial coupling constant is held fixed as the density varies, it may, even though small, be so large that the one–loop flow leaves the weak–coupling regime before the curvature effects set in. In that case, a different instability, like one towards antiferromagnetism may be stronger than the superconducting one. The details depend on the parameter regime of the model; see [9, 5, 4] for results from one–loop RG flows. It should be noted that when the curvature of the Fermi surface is very small, such as in the case where t'/t is small and the filling is near to half–filling, the higher–order terms generated by the 1PI six–point function are not suppressed by scaling factors and therefore these higher order terms should be taken into account.

1.2.3 The temperature range for Landau Fermi liquid theory

In view of the generic flow to strong coupling one may ask in what sense one can regard Landau Fermi liquid theory as a "fixed point" of the RG. Because of the Kohn–Luttinger effect, the answer is that (unless time–reversal symmetry is broken explicitly in the free system [15]) the only true weak–coupling fixed point in $d \geq 2$ is the free fermion point (coupling equal to zero), and it is unstable with respect to any weak four–fermion perturbation, be it attractive or repulsive. The true meaning of the Landau Fermi liquid "fixed point" in this connection is that, although the flow eventually leads away from the zero–coupling fixed point, it may stay in its vicinity for a very long flow time, i.e., to very low energy scales. Since temperature sets a natural energy scale at which the renormalization group flow stops, it poses a natural restriction on the maximal flow time. Thus, if the temperature is high enough, more precisely if $|g_0| \log \beta\epsilon_0 \ll 1$, the system always stays weakly coupled, and Landau Fermi liquid theory can be justified by convergent renormalized perturbation theory [2, 3, 17]. As discussed in [2, 3], this is also the physically natural definition of Fermi liquid theory: Fermi liquids can be observed only above a critical temperature where superconductivity or some other kind of symmetry breaking sets in (this temperature may be very low). Above that temperature, the k–space occupation number density does not have a discontinuity, but the Fermi surface can be identified simply by the scaling of the maximal slope in the occupation density as a function of temperature. For a Fermi liquid, the derivative of the occupation number density becomes of order β at the Fermi surface, as in the Fermi gas, where the occupation number is simply $(1+\mathrm{e}^{\beta e(p)})^{-1}$, and only the prefactor changes due to a finite wave function renormalization. For a Luttinger liquid, the anomalous scaling of the two–point function (corresponding to an infinite wave function renormalization), leads to a scaling with β^α, $\alpha < 1$. If a gap occurs, the slope does not scale with the temperature any more. This criterion also corresponds to what is done in experiments – the integrated intensity measured in photoemission experiments allows to reconstruct the occupation density if the interaction matrix elements are known [19]. Theoretically, verifying that the step in the occupation number density scales as in the Fermi gas corresponds to showing that the first two derivatives of the selfenergy are finite. This was first done in [2, 12]. Moreover, if $|g_0| \log \beta\epsilon_0 \ll 1$, the Cooper instability does not lead to large coupling constants, and therefore the Landau function describing the interaction of the quasiparticles is well–defined.

1.2.4 Larger initial coupling constants

While the above discussion was to a good part about initially weakly coupled systems, it may happen that the regime 2, where the couplings are not small but the restricted phase space justifies looking at one–loop flows, still exists. If the exact integration over the modes with energies above ϵ_1 leaves over a system with a curved Fermi surface, and if that curvature is big enough so that the corrections to the nonoverlapping part of the flow are small, then one can still use the one–loop approximation to do reliable calculations in regime 2, and thus justify Fermi liquid theory above the critical temperature of the superconducting transition. That this condition is ever fulfilled is not at all obvious but it seems possible.

Acknowledgements. It is our pleasure to thank Gianni Blatter, Joel Feldman, Walter Metzner, and Maurice Rice, for discussions.

2 The renormalization group

In this section, we present the renormalization group (RG) setup and then discuss its application to our problem. We first review the path integral representation very briefly; the RG is introduced as a transformation that leaves the generating function for the correlation functions invariant.

2.1 The path integral representation

We consider a fermionic system with creation and annihilation operators $a^+_{\mathbf{x},\sigma}$ and $a_{\mathbf{x},\sigma}$ ($\sigma = \pm$ the spin in units $\hbar/2$) which obey the canonical anticommutation relations, and with a Hamiltonian $H(a^+, a)$ that is a polynomial in the fermion operators. For instance, a Hubbard type Hamiltonian is

$$H = \sum_{\mathbf{x},\mathbf{x}',\sigma} T_{\mathbf{x}\mathbf{x}'} a^+_{\mathbf{x},\sigma} a_{\mathbf{x}',\sigma} + \sum_{\mathbf{x},\mathbf{x}'} : n_{\mathbf{x}} V(\mathbf{x},\mathbf{x}') n_{\mathbf{x}'} : \tag{2}$$

with the number operators $n_{\mathbf{x}} = a^+_{\mathbf{x},+} a_{\mathbf{x},+} + a^+_{\mathbf{x},-} a_{\mathbf{x},-}$ and the dots denoting normal ordering. The Hamiltonian is hermitian if $T_{\mathbf{x}\mathbf{y}} = \overline{T}_{\mathbf{y}\mathbf{x}}$ and if $V(\mathbf{x},\mathbf{x}') = V(\mathbf{x}',\mathbf{x})$. For a translation-invariant system, $T_{\mathbf{x}\mathbf{y}} = T(\mathbf{x}-\mathbf{y})$ and $V(\mathbf{x},\mathbf{y}) = v(\mathbf{x}-\mathbf{y})$. The usual repulsive on-site Hubbard interaction is given by $v(\mathbf{x}-\mathbf{y}) = \frac{U}{2}\delta_{\mathbf{x},\mathbf{y}}$ with $U > 0$.

The grand canonical trace $Z = \mathrm{Tr}\, \mathrm{e}^{-\beta(H-\mu N)}$ has a representation as a Grassmann functional integral

$$Z = \int \prod_{\tau,\mathbf{x},\sigma} \mathrm{d}\bar{\psi}_\sigma(\tau,\mathbf{x})\, \mathrm{d}\psi_\sigma(\tau,\mathbf{x}) \mathrm{e}^{-\mathcal{A}(\bar{\psi},\psi)} \tag{3}$$

where the Grassmann fields $\bar{\psi}$ and ψ now depend on an additional Euclidian time τ, which we choose in the interval $-\beta/2 \le \tau < \beta/2$, and which are antiperiodic in τ. That is, ψ and $\bar{\psi}$ are really defined for all $-\beta \le \tau < \beta$, they are periodic under $\tau \to \tau + 2\beta$, and satisfy

$$\begin{array}{ll} \psi_\sigma(\tau+\beta,\mathbf{x}) & = -\psi_\sigma(\tau,\mathbf{x}) \\ \bar{\psi}_\sigma(\tau+\beta,\mathbf{x}) & = -\bar{\psi}_\sigma(\tau,\mathbf{x}). \end{array} \tag{4}$$

The exponent $\mathcal{A}$ in (3) is the action corresponding to H,

$$\begin{aligned} \mathcal{A}(\bar{\psi},\psi) = \int_{-\beta/2}^{\beta/2} \mathrm{d}\tau [\quad & \textstyle\sum_{\mathbf{x},\sigma} \quad \bar{\psi}_\sigma(\tau,\mathbf{x})\, \partial_\tau \psi_\sigma(\tau,\mathbf{x}) \\ & - \quad H(\bar{\psi}(\tau),\psi(\tau))] \end{aligned} \tag{5}$$

The first summand is the analogue of the usual $p\dot{q}$ term in the Legendre transform. The second is just the Hamiltonian itself, in normal ordered form, and with $c^+_{\sigma,\mathbf{x}}$ replaced by $\bar{\psi}_\sigma(\tau,\mathbf{x})$ and $c_{\sigma,\mathbf{x}}$ replaced by $\psi_\sigma(\tau,\mathbf{x})$.

Strictly speaking, the functional integral is the limit $n \to \infty$ of a finite-dimensional Grassmann integral with discrete times $\tau_k = -\beta + \frac{k}{n}\beta$, $0 \le k < n$. At finite n, the analogue of the functional integral (3) is just a finite–dimensional Grassmann integral. Taking the limit $n \to \infty$ is a standard procedure (see, e.g. [1], [3]).

For the Hamiltonian (2), the action consists of a quadratic and a quartic term in the fields. $\mathcal{A} = \mathcal{A}_2 + \mathcal{A}_4$. The quadratic term is

$$\mathcal{A}_2(\bar{\psi},\psi) = (\bar{\psi},\ Q\psi) \tag{6}$$

where, for Grassmann fields ψ and $\bar{\eta}$, we have defined a fermionic bilinear form

$$(\bar{\eta},\ \psi) = \int \mathrm{d}\xi\ \bar{\eta}(\xi)\, \psi(\xi), \tag{7}$$

used the integral notation $\int \mathrm{d}\xi\, F(\xi) = \int_{-\beta/2}^{\beta/2} \mathrm{d}\tau \sum_{\mathbf{x},\sigma} F(\tau,\mathbf{x},\sigma)$ (with $\xi = (\tau,\mathbf{x},\sigma)$), and denoted the quadratic operator in the action by Q:

$$Q(\xi,\xi') = \delta_{\sigma\sigma'}\delta(\tau-\tau')(\delta_{\mathbf{x}\mathbf{x}'}(\partial_\tau + \mu) - T_{\mathbf{x}\mathbf{x}'}). \tag{8}$$

The bilinear form satisfies $(\bar\eta,\,\psi) = -(\psi,\bar\eta)$ because the Grassmann fields anticommute. The quartic term is

$$\begin{aligned}\mathcal{A}_4(\bar\psi,\psi) &= \int_{-\beta/2}^{\beta/2}\mathrm{d}\tau \sum_{\mathbf{x},\mathbf{x}',\sigma,\sigma'} \bar\psi_\sigma(\tau,\mathbf{x})\psi_\sigma(\tau,\mathbf{x}) \\ &\quad V(\mathbf{x},\mathbf{x}')\bar\psi_{\sigma'}(\tau,\mathbf{x}')\psi_{\sigma'}(\tau,\mathbf{x}') \end{aligned} \tag{9}$$

The action has all symmetries of the Hamiltonian. Moreover, the action and the integration measure, hence the partition function, are invariant under the transformation

$$\psi_\sigma(\tau,\mathbf{x}) \to \mathrm{i}\bar\psi_\sigma(-\tau,\mathbf{x}), \quad \bar\psi_\sigma(\tau,\mathbf{x}) \to \mathrm{i}\psi_\sigma(-\tau,\mathbf{x}). \tag{10}$$

Observables transform accordingly (similarly as when taking adjoints).

It is convenient, and natural, to combine the exponential of the quadratic part of the action and the Grassmann product measure into a Grassmann Gaussian measure

$$\mathrm{d}\mu_C(\bar\psi,\psi) = \frac{1}{\det Q}\prod_\xi \mathrm{d}\bar\psi(\xi)\,\mathrm{d}\psi(\xi)\;\mathrm{e}^{-\mathcal{A}_2(\bar\psi,\psi)} \tag{11}$$

The covariance C is the inverse of $C = Q^{-1}$. The inverse exists because of the antiperiodic boundary conditions. We also call $\mathcal{A}_4(\bar\psi,\psi) = \mathcal{V}(\bar\psi,\psi)$.

For a translation-invariant system, C can be calculated by a Fourier transform

$$C(\xi;\xi') = \frac{\delta_{\sigma\sigma'}}{\beta L^d}\sum_\omega\sum_{\mathbf{p}} \mathrm{e}^{\mathrm{i}(\tau-\tau')\omega+\mathrm{i}(\mathbf{x}-\mathbf{x}')\mathbf{p}}\,\hat C(\omega,\mathbf{p}) \tag{12}$$

with $\hat C(\omega,\mathbf{p}) = (\mathrm{i}\omega - e(\mathbf{p}))^{-1}$, and $e(\mathbf{p}) = \hat T(\mathbf{p}) - \mu$, where $\hat T$ denotes the Fourier transform of the hopping amplitude T. The summation over $\mathbf{p}$ is over the reciprocal lattice, and the sum over ω runs over all Matsubara frequencies, $\omega_n = \frac{\pi}{\beta}(2n+1)$, n integer. Thus $\hat C$ is simply the usual Matsubara propagator. The value of C at coinciding times is defined as the limit $\tau' \uparrow \tau$, as required by time ordering.

With these definitions, and with Grassmann source terms η and $\bar\eta$, the generating function for the connected Green functions is

$$W(\bar\eta,\eta) = -\log\int \mathrm{d}\mu_C(\bar\psi,\psi)\mathrm{e}^{-\mathcal{V}(\bar\psi,\psi)+(\bar\eta,\psi)+(\bar\psi,\eta)} \tag{13}$$

The Z in (3) is now $\mathrm{e}^{-W(0,0)} \det Q$. All information about the system is contained in the connected correlation functions, which are the derivatives of $\log Z(\bar\eta, \eta)$ with respect to the source terms $\bar\eta$ and η, at $\eta = \bar\eta = 0$. When $\mathcal{V} = 0$, the fermions are noninteracting, the integral for Z is Gaussian, and $W(\bar\eta, \eta) = -(\bar\eta, C\eta)$. This implies that the correlation functions are determinants of the two-point function (or covariance) C.

Thus all objects appearing in the functional integral have a natural interpretation. The Gaussian measure with covariance C describes free fields, i.e. particles with propagator C. The interaction appears in the form of an (unnormalized) Boltzmann factor. Thus our system is characterized by C and $\mathcal{V}$. The renormalization group is a transformation of C and $\mathcal{V}$ which depends on an energy scale and which leaves $Z(\bar\eta, \eta)$ invariant.

Performing a shift in the integration variables, one gets

$$\mathrm{e}^{-W(\bar\eta,\eta)} = \mathrm{e}^{(\bar\eta, C\eta)} \int \mathrm{d}\mu_C(\bar\psi, \psi) \mathrm{e}^{-\mathcal{V}(\bar\psi - C^T\bar\eta,\, \psi - C\eta)} \tag{14}$$

(here C^T is the transpose of C). The function

$$\mathcal{G}(\bar\phi, \phi) = -\log \int \mathrm{d}\mu_C(\bar\psi, \psi) \mathrm{e}^{-\mathcal{V}(\bar\psi + \bar\phi,\, \psi + \phi)} \tag{15}$$

is called the effective action. We have $W(\bar\eta, \eta) = (\bar\eta, C\eta) - \mathcal{G}(C^T\bar\eta, C\eta)$, so studying W is equivalent to studying $\mathcal{G}$.

We now streamline notation and introduce Nambu-type fields, to make the derivation of the RG equations simpler. Let $X = (\tau, \mathbf{x}, \sigma, c)$, where the charge index $c = \pm 1$ distinguishes between ψ and $\bar\psi$: with the Grassmann field $\Psi(X)$ given by

$$\Psi(\tau, \mathbf{x}, \sigma, 1) = \bar\psi_\sigma(\tau, \mathbf{x}), \qquad \Psi(\tau, \mathbf{x}, \sigma, -1) = \psi_\sigma(\tau, \mathbf{x}), \tag{16}$$

the quadratic part of the action is

$$(\bar\psi, Q\psi) = \frac{1}{2}\left((\bar\psi, \psi), \begin{pmatrix} 0 & Q \\ -Q^T & 0 \end{pmatrix} \begin{pmatrix} \bar\psi \\ \psi \end{pmatrix} \right) = \frac{1}{2}(\Psi, \mathbf{Q}\,\Psi) \tag{17}$$

with $Q^T(\xi, \xi') = Q(\xi', \xi)$. Thus $\mathbf{Q}$ satisfies $\mathbf{Q}^T = -\mathbf{Q}$. Because the source terms are

$$(\bar\eta, \psi) + (\bar\psi, \eta) = (H, \Psi) \quad \text{with } H = \begin{pmatrix} -\eta \\ \bar\eta \end{pmatrix} \tag{18}$$

we have (with $\mathbf{C} = \mathbf{Q}^{-1}$)

$$\int \mathrm{d}\mu_C(\bar{\psi}, \psi) \mathrm{e}^{(\bar{\eta},\psi)+(\bar{\psi},\eta)} = \int \mathrm{d}\mu_{\mathbf{C}}(\Psi) \mathrm{e}^{(H,\Psi)} = \mathrm{e}^{-\frac{1}{2}(H,\mathbf{C}H)}. \tag{19}$$

The exponential of the effective action $\mathcal{G}$ is the convolution

$$\mathrm{e}^{\mathcal{G}(\mathbf{C},\mathcal{V})(\Phi)} = \int \mathrm{d}\mu_{\mathbf{C}}(\Psi) \mathrm{e}^{-\mathcal{V}(\Psi+\Phi)} = \left(\mu_{\mathbf{C}} * \mathrm{e}^{\mathcal{V}}\right)(\Phi) \tag{20}$$

We shall not require $\mathbf{Q}$ to be of the form in (17) but only that it is antisymmetric. Thus it may have diagonal terms, which correspond to non–charge invariant terms, such as a superconducting gap.

2.2 RG flow: general setup

The semigroup law of the RG is based on the addition principle for Gaussian fields: if $\mathbf{C} = \mathbf{C}_1 + \mathbf{C}_2$, then Ψ splits into two independent fields, $\Psi = \Psi_1 + \Psi_2$, where Ψ_1 has covariance $\mathbf{C}_1$ and Ψ_2 has covariance $\mathbf{C}_2$. The corresponding Gaussian measure factorizes, so that

$$\begin{aligned} & \int \mathrm{d}\mu_{\mathbf{C}}(\Psi) \mathrm{e}^{-\mathcal{V}(\Psi+\Phi)} \\ = & \int \mathrm{d}\mu_{\mathbf{C}_1}(\Psi_1) \int \mathrm{d}\mu_{\mathbf{C}_2}(\Psi_2) \mathrm{e}^{-\mathcal{V}(\Psi_1+\Psi_2+\Phi)}. \end{aligned} \tag{21}$$

When one starts out at weak coupling, it is natural to label energy scales according to the kinetic energy. In the following, we shall use the decomposition $\mathbf{C} = \mathbf{D}_s + \mathbf{C}_s$ with a scale parameter $s \geq 0$ corresponding to an energy scale $\epsilon_s = \epsilon_0 \mathrm{e}^{-s}$. The propagator $\mathbf{C}_s$ corresponds to fields with energy $e(\mathbf{p})$ above ϵ_s, and $\mathbf{D}_s$ to the fields with energy below ϵ_s. The effective action where the fields with propagator $\mathbf{C}_s$ have been integrated out is $\mathcal{G}_{\epsilon_s} = \mathcal{G}(\mathbf{C}_s, \mathcal{V})$. In the limit $s \to \infty$, where the energy scale ϵ_s vanishes, $\mathbf{D}_s$ is identically zero and everything has been integrated out. Inserting this decomposition and using (21), we get for the effective action $\mathcal{G}_0 = \mathcal{G}(\mathbf{C}, \mathcal{V})$, in which all fields have been integrated over,

$$\begin{aligned} \mathrm{e}^{-\mathcal{G}_0(\Phi)} &= \int \mathrm{d}\mu_{\mathbf{D_s}+\mathbf{C_s}}(\Psi) \mathrm{e}^{-\mathcal{V}(\Psi+\Phi)} \\ &= \int \mathrm{d}\mu_{\mathbf{D_s}}(\Psi') \mathrm{e}^{-\mathcal{G}_{\epsilon_s}(\Psi'+\Phi)}. \end{aligned} \tag{22}$$

Eq. (22) is the semigroup law of the RG; it can also be written in the form

$$\mathcal{G}(\mathbf{D}_s + \mathbf{C}_s, \mathcal{V}) = \mathcal{G}(\mathbf{D}_s, \mathcal{G}(\mathbf{C}_s, \mathcal{V})). \tag{23}$$

The semigroup law imples that the system $(\mathbf{C}, \mathcal{V})$ that we want to study is exactly equivalent to the system $(\mathbf{D}_s, \mathcal{G}(\mathbf{C}_s, \mathcal{V}))$ with propagator $\mathbf{D}_s$ and effective interaction $\mathcal{G}(\mathbf{C}_s, \mathcal{V})$. It also shows that the effective action $\mathcal{G}(\mathbf{C}_s, \mathcal{V})$ generates the connected, amputated correlation functions of the model with covariance $\mathbf{C}_s$. In our application, $\mathbf{C}_s$ is a covariance with an infrared cutoff ϵ_s, and $\mathbf{D}_s$ is supported on a smaller momentum space because it is nonzero only for fields with energies smaller than ϵ_s.

Set up in this way, the renormalization group is simply a symmetry of the generating functional $\mathcal{G}(\mathbf{C}, \mathcal{V})$, which contains all information of the model. The RG differential equation expresses this statement in differential form, namely that $\mathcal{G}(\mathbf{C}, \mathcal{V})$ is independent of s. Thus the RGDE is the equation $\frac{\partial}{\partial s}\mathcal{G}(\mathbf{C}, \mathcal{V}) = 0$, evaluated by inserting the right hand side of (22).

The exact symmetry comes at a price: $\mathcal{G}$ is a more complicated function than the original interaction. It is an infinite power series

$$\begin{aligned} \mathcal{G}_{\epsilon_s}(\psi) &= \sum_{\sigma} \int \mathrm{d}p\, \bar{\psi}_\sigma(p) \hat{B}_{\epsilon_s}(p) \psi_\sigma(p) \\ + &\sum_{\sigma_1,\ldots,\sigma_4} \int \mathrm{d}p_1 \ldots \mathrm{d}p_4 \bar{\psi}_{\sigma_1}(p_1) \bar{\psi}_{\sigma_2}(p_2) \psi_{\sigma_3}(p_3) \psi_{\sigma_4}(p_4) \\ &\delta(p_1 + p_2 - p_3 - p_4) \hat{F}^{(\epsilon_s)}_{\sigma_1,\ldots,\sigma_4}(p_1, p_2, p_3) \\ + &\quad \text{terms of order } \psi^6, \quad \psi^8, \ldots \end{aligned} \tag{24}$$

in the fields. We shall mainly be concerned with the quadratic and quartic terms, which have the significance of the selfenergy correction and the effective interaction; however, one should be aware that the interactions corresponding to the terms of order ψ^6 or higher are always there and that the convergence of the infinite series is a nontrivial problem (it has been shown to converge in two dimensions; see [13]). The selfenergy term B produces, among other things, a Fermi surface shift, which has to be kept track of in the flow. This has been done for general Fermi surfaces [11].

In the following, we derive a differential equation for the one-particle-irreducible (1PI) functions, from which we can obtain the above functions B

and F. This derivation is similar to the one used in [2], except that we do not use Wick ordering here and give the equations for the 1PI functions. The Wick-ordered equation of [2] is used in [5]. Yet another form (the original one of Polchinski) is used in [4]. In an exact treatment, all these approaches are equivalent because they give the same correlation functions. In approximations obtained by truncations of the equations, the question of equivalence becomes nontrivial. From the technical point of view, there are a number of differences, on which we shall comment later.

To summarize, the RG expresses the generating function of our field theory in the form

$$\int \mathrm{d}\mu_{\mathbf{C}}(\Psi)\mathrm{e}^{-\mathcal{V}(\Psi+\Phi)} = \int \mathrm{d}\mu_{\tilde{\mathbf{D}}_s}(\Psi)\mathrm{e}^{-\tilde{\mathcal{V}}_s(\Psi+\Phi)}\,. \tag{25}$$

In the way derived above, $\tilde{\mathbf{D}}_s = \mathbf{D}_s$ and $\tilde{\mathcal{V}}_s = \mathcal{G}(\mathbf{C}_s, \mathcal{V})$, but other choices are possible: $\mathcal{G}(\mathbf{C}_s, \mathcal{V})$ will in general contain a selfenergy term (quadratic in the fields) which we shall absorb in the new propagator $\tilde{\mathbf{D}}_s$ later to deal with the flow of the Fermi surface and the superconducting gap.

Note that because the generating function is invariant under the decomposition, we shall be able to stop the flow at any scale we want; this will be necessary because our flows generically lead to strong coupling.

3 The RG for the 1PI functions

In the following, we derive the RG for the 1PI functions. For the moment, we shall not need all the details of the setup in terms of scale parameters; we shall only use that $\mathbf{C}_s$ depends on a parameter s. Recall that studying W (given in (13)) is equivalent to studying $\mathcal{G}$. The 1PI functions are generated by the first Legendre transform Γ of W.

3.1 The generating function of the 1PI vertices

In a general theory with fermionic fields, the fields $\psi(X)$ and $\bar{\psi}(X)$ are labelled by an index X which comprises spacetime, spin, flavour, and possible other indices. We collect ψ and $\bar{\psi}$ into a single vector $\Psi = (\bar{\psi}, \psi)$. We also use the notation $(A, B) = \int \mathrm{d}X A(X)B(X)$, where $\int \mathrm{d}X$ stands for summation over the discrete indices and integrals over the continuous ones. In the

Hubbard model, the standard functional integral representation (see, e.g. [3], Section 4.2) gives $X = (\tau, \mathbf{x}, \sigma, c)$, where $\mathbf{x}$ is the position, $\sigma = \pm$ the third component of the spin, $-\beta/2 \leq \tau < \beta/2$ the usual Euclidian time used to convert the grand canonical trace to a functional integral over the Grassmann fields Ψ, and where the charge index $c = \pm$ distinguishes between the components ψ and $\bar{\psi}$ of Ψ.

The generating function for the connected non–amputated Green functions is defined by

$$\mathrm{e}^{-W(H)} = \int \mathrm{d}\mu_{\mathbf{C}}(\Psi) \mathrm{e}^{-\mathcal{V}(\Psi)+(H,\Psi)} \,. \tag{26}$$

Here the Gaussian integral is given by an invertible operator $\mathbf{Q}$ with integral kernel $\mathbf{Q}(X, X')$. Because of the Grassmann nature of the fields, $\mathbf{Q}$ is antisymmetric, i.e. $\mathbf{Q}(X', X) = -\mathbf{Q}(X, X')$. The covariance $\mathbf{C}$ is $\mathbf{C} = \mathbf{Q}^{-1}$, and $\mathrm{d}\mu_{\mathbf{C}}(\Psi) = (\det \mathbf{Q})^{-1} \mathrm{e}^{-\frac{1}{2}(\Psi, \mathbf{Q}\Psi)} D\bar{\psi} D\psi$, with the notation (A, B) as defined above. A general antisymmetric $\mathbf{Q}$ includes from the start the possibility of non–charge invariant terms of type $\psi(X)\psi(X')$; charge invariance corresponds to a matrix $\mathbf{Q}$ of the form

$$(\mathbf{Q}(\xi, \xi'))_{cc'} = \begin{pmatrix} 0 & Q(\xi, \xi' \\ -Q(\xi', \xi) & 0 \end{pmatrix} . \tag{27}$$

In the Hubbard model,

$$Q(\xi, \xi') = \delta_{\sigma\sigma'} \delta(\tau - \tau') (\delta_{\mathbf{x}\mathbf{x}'} (\partial_\tau + \mu) - T_{\mathbf{x}\mathbf{x}'}), \tag{28}$$

where T denotes the hopping matrix. $\mathcal{V}$ is the interaction term written in terms of the fields Ψ (for details, see e.g. [3]).

The source term H is another Grassmann vector; if $H = \binom{-\eta}{\bar{\eta}}$, then (H, Ψ) is the usual combination $(\bar{\eta}, \psi) + (\bar{\psi}, \eta)$.

If W has a nondegenerate quadratic part, the map $H \mapsto \Phi_{cl}(H)$, with

$$\Phi_{cl}(H)(X) = \frac{\delta}{\delta H(X)} W(H), \tag{29}$$

can be inverted (this is the case in our fermionic models because the Grassmann variables are nilpotent and the covariance $\mathbf{C}$ is nondegenerate at positive temperature; in bosonic models, the map would not be invertible if

symmetry breaking occurs; the Legendre transform is then defined by a variational equation; see, e.g. [7]). Denote the inverse map by $\phi \mapsto h(\phi)$ (h is an odd element of the Grassmann algebra generated by ϕ), so that

$$\Phi_{cl}(h(\phi)) = \frac{\delta W}{\delta H}(h(\phi)) = \phi\,. \tag{30}$$

Taking a derivative with respect to ϕ gives

$$\int \mathrm{d}Z \frac{\delta h(\phi)(Z)}{\delta \phi(Y)} \left(\frac{\delta^2 W}{\delta H(Z)\delta H(X)}\right)(h(\phi)) = \delta(X,Y)\,. \tag{31}$$

The first Legendre transform of W is

$$\Gamma(\phi) = W\left(h(\phi)\right) - (h(\phi), \phi) \tag{32}$$

(with the last term a bilinear form as above); it generates the 1PI correlation functions. We have $\frac{\delta \Gamma}{\delta \phi} = h(\phi)$ and thus by (31) (as operators)

$$\left(\frac{\delta^2 \Gamma}{\delta \phi^2}\right)(\phi) = \left[\frac{\delta^2 W}{\delta H^2}(h(\phi))\right]^{-1} \tag{33}$$

For free particles ($\mathcal{V} = 0$), $W = \frac{1}{2}(H, \mathbf{C}H)$, so $\delta W/\delta H = \mathbf{C}H$, hence $h(\phi) = \mathbf{C}^{-1}\phi$, and $\Gamma(\phi) = \frac{1}{2}(\phi, \mathbf{C}^{-1}\phi)$. In first order, the four–fermion interaction term in Γ is just the original interaction $\mathcal{V}$.

3.2 The RG differential equation for Γ

If W depends on a parameter s, then Γ and h also depend on s. By (30),

$$\frac{\mathrm{d}}{\mathrm{d}s} W_s(h_s(\phi)) = \frac{\partial W_s}{\partial s}(h_s(\phi)) + (\dot{h}_s(\phi), \phi) \tag{34}$$

(where the dot denotes the derivative with respect to s), so (32) implies

$$\dot{\Gamma}_s(\phi) = \dot{W}_s(h_s(\phi)). \tag{35}$$

We now assume that the s–dependence of W_s is given as follows. In (26), $\mathcal{V}$ remains independent of s, but $\mathbf{C}$ gets replaced by $\mathbf{C}_s = \mathbf{Q}_s^{-1}$, where $\mathbf{Q}_s$ now depends on s. Then the derivative $\partial/\partial s$ can act only on $\mathrm{d}\mu_{\mathbf{C}_s}$, that is, on the

normalization factor or on the exponent. In the former case, it just produces a constant term; in the latter it brings down $(\Psi, \dot{\mathbf{Q}}_s\Psi)$ in the integral. Using

$$(\Psi, \dot{\mathbf{Q}}_s\Psi)\, \mathrm{e}^{(H,\Psi)} = \left(\frac{\delta}{\delta H}, \dot{\mathbf{Q}}_s \frac{\delta}{\delta H}\right)\, \mathrm{e}^{(H,\Psi)} \tag{36}$$

we can reexpress everything in terms of $W_s(H)$ and get

$$\begin{aligned} \dot{W}_s(H) &= \frac{1}{2}\,\mathrm{Tr}\,(\mathbf{C}_s\dot{\mathbf{Q}}_s) + \frac{1}{2}\left(\frac{\delta}{\delta H}, \dot{\mathbf{Q}}_s\frac{\delta}{\delta H}\right) W_s(H) \\ &+ \frac{1}{2}\left(\frac{\delta W_s}{\delta H}, \dot{\mathbf{Q}}_s\frac{\delta W_s}{\delta H}\right), \end{aligned} \tag{37}$$

and $\mathrm{Tr}\,(AB) = \int \mathrm{d}X\,\mathrm{d}Y A(X,Y)B(Y,X)$. This is an equation similar to Polchinski's equation [6], but with $\mathbf{Q}_s$ instead of $\mathbf{C}_s$ in the Laplacian because the Green functions generated by W are not amputated. By (35), (30), and (33), the differential equation for $\Gamma(s)$ is

$$\begin{aligned} \dot{\Gamma}(s|\phi) &= \frac{1}{2}\,\mathrm{Tr}\,(\mathbf{C}_s\dot{\mathbf{Q}}_s) + \frac{1}{2}(\phi, \dot{\mathbf{Q}}_s\phi) \\ &+ \frac{1}{2}\,\mathrm{Tr}\,\left[\dot{\mathbf{Q}}_s\left(\frac{\delta^2\Gamma(s|\phi)}{\delta\phi^2}\right)^{-1}\right]. \end{aligned} \tag{38}$$

This is a nonpolynomial equation for Γ, but the inverse contains a second derivative, which produces a field–independent term coming from the quadratic term in Γ. Thus the equation makes sense in an expansion in the fields.

3.3 Expansion in the fields

In this section we derive the equation for the scale–dependent 1PI m–point functions $\gamma_m(s)$, by expanding $\Gamma(s|\phi)$ in the fields. Readers that only want to see the result can skip to the next subsection.

The 1PI m-point vertex functions $\gamma_m(s|X_1, \ldots, X_m)$ are the coefficients in an expansion of Γ as a power series in the fields:

$$\Gamma(s|\phi) = \sum_{m\geq 0} \gamma^{(m)}(s|\phi) \tag{39}$$

with

$$\gamma^{(m)}(s|\phi) = \frac{1}{m!}\int \mathrm{d}^m \underline{X}\, \gamma_m(s|\underline{X})\, \phi^m(\underline{X})\,. \tag{40}$$

Here we used the notations $\underline{X} = (X_1, ..., X_m)$ and $\phi^m(\underline{X}) = \phi(X_1)...\phi(X_m)$. We choose the function $\gamma_m(s|\underline{X})$ totally antisymmetric: $\mathcal{A}_m\gamma_m = \gamma_m$, where $(\mathcal{A}_m F)(X_1, \ldots, X_m) = \frac{1}{m!}\sum_\pi \epsilon(\pi) F(X_{\pi(1)}, \ldots, X_{\pi(m)})$ is the antisymmetrization operator (π is summed over all permutations of $1, \ldots, m$ and $\epsilon(\pi)$ denotes the sign of π). Any part of γ_m that is not antisymmetric would cancel out in (40) because of the antisymmetry properties of the Grassmann variables. Also, we shall later compare coefficients; this is allowed only when totally antisymmetric functions are used. The $\gamma_m(s|\underline{X})$ are the 1PI vertex functions. Similarly, we have the expansion

$$\frac{\delta}{\delta\phi(X)}\frac{\delta}{\delta\phi(Y)}\Gamma(s|\phi) = \sum_{m\geq 0}\tilde{\gamma}^{(m)}(s|X, Y; \phi)\,. \tag{41}$$

By the antisymmetry of γ_m, two derivatives applied to $\gamma^{(m+2)}$ give a factor $(m+2)(m+1)$, which combines with the $1/(m+2)!$ to $1/m!$ (this is the reason for the convention of putting the prefactor $\frac{1}{m!}$ in (40)). Thus

$$\tilde{\gamma}^{(m)}(s|X, Y; \phi) = \frac{1}{m!}\int \mathrm{d}^m \underline{X}'\gamma_{m+2}(s|X, Y, \underline{X}')\phi^m(\underline{X}')\,. \tag{42}$$

In particular, $\tilde{\gamma}^{(0)}$ is independent of ϕ:

$$\tilde{\gamma}^{(0)}(s|X, Y; \phi) = \gamma_2(s|X, Y)\,. \tag{43}$$

Therefore

$$\frac{\delta^2\Gamma(s|\phi)}{\delta\phi(X)\delta\phi(Y)} = \gamma_2(s|X, Y) + \tilde{\Gamma}(s|X, Y; \phi) \tag{44}$$

with

$$\tilde{\Gamma}(s|X, Y; \phi) = \sum_{m\geq 2}\tilde{\gamma}^{(m)}(s|X, Y; \phi)\,. \tag{45}$$

It is natural to think of $\gamma_2(s|X, Y)$ and of $\tilde{\Gamma}(s|X, Y; \phi)$ as integral kernels of operators γ_2 and $\tilde{\Gamma}(s|\phi)$. By the relation (33) at $\phi = 0$,

$$\mathbf{G}_s = \gamma_2(s)^{-1} \tag{46}$$

is the full two-point function. As an equation between operators, we thus have

$$\frac{\delta^2\Gamma}{\delta\phi^2}(s|\phi) = \gamma_2(1 + \mathbf{G}_s\tilde{\Gamma}(s|\phi))\,, \tag{47}$$

so the differential equation for Γ now reads

$$\begin{aligned}\dot{\Gamma}(s|\phi) &= \frac{1}{2}\,\mathrm{Tr}\,(\mathbf{C}_s\dot{\mathbf{Q}}_s) + \frac{1}{2}(\phi,\dot{\mathbf{Q}}_s\phi) \\ &+ \frac{1}{2}\,\mathrm{Tr}\,\left[\mathbf{G}_s\dot{\mathbf{Q}}_s(1+\mathbf{G}_s\tilde{\Gamma}(s|\phi))^{-1}\right] \end{aligned} \tag{48}$$

This equation is nonpolynomial in Γ because of the irreducibility condition. We get the equations for the γ_m by expanding $(1+\mathbf{G}_s\tilde{\Gamma}(s|\phi))^{-1}$ in a geometric series

$$\begin{aligned}&\mathrm{Tr}\,\left[\mathbf{G}_s\dot{\mathbf{Q}}_s(1+\mathbf{G}_s\tilde{\Gamma}(s\mid\phi))^{-1}\right] \\ = \;& \mathrm{Tr}\,(\mathbf{G}_s\dot{\mathbf{Q}}_s) - \mathrm{Tr}\,(\mathbf{G}_s\dot{\mathbf{Q}}_s\mathbf{G}_s\tilde{\Gamma}(s|\phi)) \\ + \;& \sum_{p\geq 2}(-1)^p\,\mathrm{Tr}\,\left[\mathbf{G}_s\dot{\mathbf{Q}}_s(\mathbf{G}_s\tilde{\Gamma}(s|\phi))^p\right] \end{aligned} \tag{49}$$

The first term is a constant, which corresponds to a vacuum energy and is not interesting for our purposes because it drops out in all correlation functions. The term linear in $\tilde{\Gamma}$ generates contractions with single lines; its lowest order in ϕ is quadratic in ϕ and therefore generates selfenergy corrections. The graphical interpretation of the terms with $p \geq 2$ is also straightforward: The p'th oder term $(\mathbf{G}_s\tilde{\Gamma}(s,\phi))^p\mathbf{G}_s$ is a linear tree with p vertices. Taking the trace with $\dot{\mathbf{Q}}_s$ forms a loop. Thus in the graphical expansion, only 1PI graphs contribute to Γ.

We define the single-scale propagator as

$$\mathbf{S}_s = -\mathbf{G}_s\dot{\mathbf{Q}}_s\mathbf{G}_s. \tag{50}$$

The $\tilde{\gamma}^{(m)}$ defined in (42) are homogeneous of degree m in ϕ; inserting (39) on the left hand side and (45) on the right hand side of (48), we get a system of equations for the $\gamma^{(m)}$. For $m \leq 6$ the equations are

$$\dot{\gamma}^{(2)}(s|\phi) = \frac{1}{2}(\phi,\dot{\mathbf{Q}}_s\phi) + \frac{1}{2}\,\mathrm{Tr}\,(\mathbf{S}_s\tilde{\gamma}^{(2)}(s|\phi)) \tag{51}$$

$$\begin{aligned}\dot{\gamma}^{(4)}(s|\phi) &= \frac{1}{2}\,\mathrm{Tr}\,[\mathbf{S}_s\tilde{\gamma}^{(4)}(s\mid\phi)] \\ &- \frac{1}{2}\,\mathrm{Tr}\,[\mathbf{S}_s\tilde{\gamma}^{(2)}(s\mid\phi)\mathbf{G}_s\tilde{\gamma}^{(2)}(s|\phi)]\end{aligned} \tag{52}$$

$$\dot{\gamma}^{(6)}(s|\phi) = \frac{1}{2}\,\mathrm{Tr}\,[\mathbf{S}_s\tilde{\gamma}^{(6)}(s|\phi)]$$

$$
\begin{aligned}
&- \frac{1}{2}\,\mathrm{Tr}\,[\mathbf{S}_s(\tilde\gamma^{(4)}\mathbf{G}_s\tilde\gamma^{(2)}+\tilde\gamma^{(2)}\mathbf{G}_s\tilde\gamma^{(4)}]\\
&+\frac{1}{2}\,\mathrm{Tr}\,[\mathbf{S}_s\tilde\gamma^{(2)}\mathbf{G}_s\tilde\gamma^{(2)}\mathbf{G}_s\tilde\gamma^{(2)}]
\end{aligned}
\tag{53}
$$

The last term contributing to $\dot\gamma^{(4)}$ is

$$
\mathrm{Tr}\,(\mathbf{S}_s\tilde\gamma^{(2)}\mathbf{G}_s\tilde\gamma^{(2)}) = \int \mathrm{d}X_1..\,\mathrm{d}X_4\,\phi(X_1)..\phi(X_4)\mathcal{T}_4 \tag{54}
$$

with

$$
\begin{aligned}
\mathcal{T}_4 \;=\; & \int \mathrm{d}Y_1\ldots\,\mathrm{d}Y_4\;\mathbf{S}_s(Y_1,Y_2)\;\mathbf{G}_s(Y_3,Y_4)\\
& \frac{1}{2}\gamma_4(s|X_1,X_2,Y_2,Y_3)\;\frac{1}{2}\gamma_4(s|X_3,X_4,Y_4,Y_1)\,.
\end{aligned}
\tag{55}
$$

Thus comparing coefficients of $\phi(X_1)\ldots\phi(X_4)$ gives, with $\underline{X}=(X_1,\ldots,X_4)$,

$$
\begin{aligned}
\dot\gamma_4(s|\underline{X}) \;=\; & \frac{1}{2}\int \mathrm{d}Y_1\,\mathrm{d}Y_2\,\gamma_6(s|\underline{X},Y_1,Y_2)\mathbf{S}_s(Y_2,Y_1)\\
& - \frac{4!}{2}\mathcal{A}_4\mathcal{T}_4(\underline{X})
\end{aligned}
\tag{56}
$$

(note that $\mathcal{T}_4$, as given in (55), is not in antisymmetrized form, so $\mathcal{A}_4\mathcal{T}_4$ appears in (56) because coefficients can only be compared if they are all antisymmetric). To calculate $\mathcal{A}_4\mathcal{T}_4$, we have to antisymmetrize the product $A_{12}B_{34}$ where $A_{kl}=\gamma_4(s|X_k,X_l,Y_1,Y_3)$ and $B_{kl}=\gamma_4(s|X_k,X_l,Y_4,Y_1)$. The antisymmetrization amounts to getting

$$
P_4(\underline{X},\underline{Y}) = \frac{1}{4}\sum_{\pi\in\mathcal{S}_4}\epsilon(\pi)A_{\pi(1)\pi(2)}B_{\pi(3)\pi(4)}\,. \tag{57}
$$

By definition, $A_{kl}=-A_{lk}$ and $B_{kl}=-B_{lk}$. The extra factor $\frac{1}{4}$ comes from the two factors $\frac{1}{2}$ in (55). We block the sum over π into those for which $\pi(\{1,2\})=\{a,b\}$ is fixed. This gives $\binom{4}{2}=6$ terms. All of them are similar, so we discuss the case $a=1$, $b=2$. There are four permutations that leave $\{1,2\}$ fixed, namely the identity, the transposition τ_{12} of 1 and 2, τ_{34} and $\tau_{12}\circ\tau_{34}$. Taking into account the signs, we have

$$
\begin{aligned}
& \sum_{\pi:\pi(\{1,2\})=\{1,2\}} \epsilon(\pi)A_{\pi(1)\pi(2)}B_{\pi(3)\pi(4)}\\
= \;& (A_{12}-A_{21})(B_{34}-B_{43}) = 4A_{12}B_{34}
\end{aligned}
\tag{58}
$$

by the antisymmetry of A and B. All other cases of $\pi(\{1,2\})$ are similar; there is a global sign from permuting 1 to a and 2 to b which for $a < b$ is $(-1)^{a+b-1}$. Thus

$$\begin{aligned} P_4(\underline{X},\underline{Y}) &= A_{12}B_{34} - A_{13}B_{24} + A_{14}B_{23} \\ &+ A_{23}B_{14} - A_{24}B_{13} + A_{34}B_{12}. \end{aligned} \tag{59}$$

Inserting the definitions of A and B and relabelling the integration variables, we can combine the terms containing A_{12} and B_{12}, A_{13} and B_{13}, etc.

3.4 The RGDE for the 1PI functions; Truncation

Denote $\underline{Y} = (Y_1, \ldots, Y_4)$,

$$\begin{aligned} \mathbf{L}(\underline{Y}) &= \mathbf{S}_s(Y_1,Y_2)\mathbf{G}_s(Y_3,Y_4) \\ &+ \mathbf{S}_s(Y_3,Y_4)\mathbf{G}_s(Y_1,Y_2), \end{aligned} \tag{60}$$

with $\mathbf{S}_s$ as in (50) and $\mathbf{G}_s$ as in (46) and

$$\begin{aligned} & \mathbf{B}_s(\underline{X},\underline{Y}) = \\ & \gamma_4(s|X_1,X_2,Y_2,Y_3)\gamma_4(s \mid Y_4,Y_1,X_3,X_4) \\ - \quad & \gamma_4(s|X_1,X_3,Y_2,Y_3)\gamma_4(s \mid Y_4,Y_1,X_2,X_4) \\ + \quad & \gamma_4(s|X_1,X_4,Y_2,Y_3)\gamma_4(s \mid Y_4,Y_1,X_2,X_3). \end{aligned} \tag{61}$$

The differential equation for the 1PI four–point function γ_4 is

$$\begin{aligned} \dot\gamma_4(s|\underline{X}) &= \frac{1}{2}\int \mathrm{d}Y_1\,\mathrm{d}Y_2 \gamma_6(s|\underline{X},Y_1,Y_2)\mathbf{S}_s(Y_2,Y_1) \\ &- \frac{1}{2}\int \mathrm{d}^4\,\underline{Y}\mathbf{L}(\underline{Y})\mathbf{B}_s(\underline{X},\underline{Y}). \end{aligned} \tag{62}$$

By (42) and (43), the equation for the 1PI two–point function γ_2 becomes

$$\begin{aligned} & \dot\gamma_2(s|X_1,X_2) = \dot{\mathbf{Q}}_s(X_1,X_2) \\ + \quad & \frac{1}{2}\int \mathrm{d}X_3\,\mathrm{d}X_4\,\mathbf{S}_s(X_4,X_3)\,\gamma_4(s|X_1,X_2,X_3,X_4)\,. \end{aligned} \tag{63}$$

Eq. (62) and (63) are the first two equations in the infinite system of RG equations (labelled by m). Note that they do not form a closed system

because γ_6 enters in (62). This behaviour continues to all m: the right hand side of the equation for $\dot{\gamma}_m$ contains γ_{m+2}.

A way to close the system of equations for the 1PI four–point function γ_4 and the selfenergy is to drop the 1PI six-point vertex from (62). This truncation is equivalent to setting all 1PI functions with $m \geq 6$ external legs zero, so that the connected (non–1PI) m-point functions with $m \geq 6$ are given by tree graphs made of the four–legged vertices and the approximation to the full propagator provided by the solution of the differential equations. The four–point and two–point differential equations are given in terms of one–loop diagrams.

Note, however, that even the untruncated system of differential equations only contains one–loop terms in every equation. This is so because in the differential formulation, only one differentiated propagator appears in the equation (and there are no tree terms in an equation for 1PI functions). Of course, this does not imply that only one–loop graphs appear in the solution; the full RG produces, after all, the full Green functions. The perturbation expansion is obtained by integrating the differential equation from 0 to s and then iterating the thus obtained integral equation until only bare vertices appear. Upon iteration, graphs with an arbitrary number of loops are generated, and if one uses the untruncated equations, all graphs are generated.

The truncated equations amount to a summation of part of the diagrams, but these diagrams also contain two–loop graphs, in particular two–loop graphs corresponding to the selfenergy.

The RG strategy does not necessarily aim at taking into account as many graphs as possible but to single out the important ones by their scaling behaviour. We discuss in Sect. 5 in which cases, and in which energy regimes, the scaling behaviour justifies a truncation of the RGDE.

The initial condition for γ_4 is the bare interaction. To renormalize the Fermi surface correctly, one also needs to take into account a Fermi surface counter term (see [14, 11, 12]).

4 Consequences of symmetries

The derivation of equations (62) and (63) did not require any symmetries, so these equations are also valid when symmetries are broken. In our systems, this means that they also hold in presence of a superconducting gap or magnetic ordering or translational symmetry breaking. In two dimensions, continuous symmetry breaking is impossible at any positive temperature by the Mermin–Wagner theorem. A noninvariance of the effective action leads immediately to long range order, hence mean–field type results. In order to compare competing instabilities, we therefore first assume that all continuous symmetries of the action remain unbroken. This leads to further simplifications in the differential equations, which we now successively discuss.

4.1 Charge invariance

Recalling that $X = (\xi, c)$ where ξ consists of space, time, and spin indices, and where $c = \pm$ is the charge index, this means that $\mathbf{S}_s((\xi, c), (\xi', c'))$ and $\mathbf{G}_s((\xi, c), (\xi', c'))$ are nonzero only if $c \neq c'$, and that $\gamma_4(s|X_1, \ldots, X_4) \neq 0$ only if two of the charge indices are $+$ and two are $-$. Because γ_4 is antisymmetric in all arguments, it is then determined by $f(s|\xi_1, \ldots, \xi_4) = \gamma_4(s|(\xi_1, +), (\xi_2, +), (\xi_3, -), (\xi_4, -))$. Also, f inherits the antisymmetry under exchange of ξ_1 and ξ_2 and that under exchange of ξ_3 and ξ_4.

We now rewrite (62) as an equation for f. The only thing to do is to arrange the internal charge index sums. For instance, the first term on the right hand side of (62) becomes

$$\begin{aligned} \sum_{c_1,\ldots,c_4} \int \mathrm{d}\eta_1, \ldots \mathrm{d}\eta_4 \, \mathbf{L}((\eta_1, c_1), \ldots, (\eta_4, c_4)) \\ \gamma_4(s \mid (\xi_1, +), (\xi_2, +), (\eta_2, c_2), (\eta_3, c_3)) \\ \gamma_4(s \mid (\eta_4, c_4), (\eta_1, c_1), (\xi_3, -), (\xi_4, -)) \,. \end{aligned} \tag{64}$$

The above-mentioned restrictions posed by γ_4 imply that the only nonzero term is $c_1 = c_4 = +$ and $c_2 = c_3 = -$, so this is

$$\begin{aligned} \int \mathrm{d}\eta_1 \ldots \mathrm{d}\eta_4 \, \mathbf{L}((\eta_1, +), (\eta_2, -), (\eta_3, -), (\eta_4, +)) \\ f(s|\xi_1, \xi_2, \eta_2, \eta_3) f(s|\eta_4, \eta_1, \xi_3, \xi_4) \,. \end{aligned} \tag{65}$$

The second and third term are similarly expressed in terms of f, using the antisymmetry of γ_4 and the charge-invariance properties. Using that

$$\begin{array}{ll} \mathbf{G}_s((\xi,+),(\xi',-)) & = -G_2(\xi',\xi) \\ \mathbf{G}_s((\xi,-),(\xi',+)) & = G_2(\xi,\xi') \end{array} \tag{66}$$

and similarly for $\mathbf{S}_s$, we get

$$\begin{array}{rcl} \dot{f}(s|\xi_1,\xi_2,\xi_3,\xi_4) & = & \Phi_{\rm pp}(s|\xi_1,\xi_2,\xi_3,\xi_4) \\ & + & \Phi_{\rm ph}(s|\xi_1,\xi_2,\xi_3,\xi_4) \\ & - & \Phi_{\rm ph}(s|\xi_1,\xi_2,\xi_4,\xi_3) \end{array} \tag{67}$$

with

$$\begin{aligned} \Phi_{\rm pp}(s|\xi_1,..,\xi_4) = \frac{1}{2}\int \mathrm{d}\eta_1..\,\mathrm{d}\eta_4\, L(\eta_2,\eta_1,\eta_3,\eta_4) \\ f(s|\xi_1,\xi_2,\eta_2,\eta_3) f(s \mid \eta_4,\eta_1,\xi_3,\xi_4) \end{aligned} \tag{68}$$

and

$$\begin{aligned} \Phi_{\rm ph}(s|\xi_1,..,\xi_4) = -\int \mathrm{d}\eta_1..\,\mathrm{d}\eta_4\, L(\eta_1,\eta_2,\eta_3,\eta_4) \\ f(s|\eta_4,\xi_2,\xi_3,\eta_1) f(s|\xi_1,\eta_2,\eta_3,\xi_4) \end{aligned} \tag{69}$$

and where

$$\begin{array}{rcl} L(\eta_1,..,\eta_4) & = & S_s(\eta_1,\eta_2) G_s(\eta_3,\eta_4) \\ & + & G_s(\eta_1,\eta_2) S_s(\eta_3,\eta_4)\,. \end{array} \tag{70}$$

There is no $\frac{1}{2}$ in $\Phi_{\rm ph}$ because there are twice as many terms in the sum over intermediate charge indices c_i in the $\Phi_{\rm ph}$ as in $\Phi_{\rm pp}$. The function $\Phi_{\rm pp}$ is antisymmetric under exchange of (ξ_1,ξ_2) and of (ξ_3,ξ_4) because f has these properties. The function $\Phi_{\rm ph}$ is not, but the difference appearing in (67) is antisymmetric.

Equation (67) has the graphical representation shown in Figure 2. The internal lines in these graphs correspond to "full" propagators G_s, and to single scale propagators S_s, respectively. The inverse of G_s is $g_2(s|\xi_1,\xi_2) = \gamma_2(s|(\xi_1,+),(\xi_2,-))$ and satisfies

$$\begin{array}{rcl} \dot{g}_2(s|\xi_1,\xi_2) & = & \dot{Q}_s(\xi_1,\xi_2) \\ & - & \displaystyle\int \mathrm{d}\xi_1\,\mathrm{d}\xi_2\; S_s(\xi_4,\xi_3)\; f(s|\xi_1,\xi_3,\xi_4,\xi_2). \end{array} \tag{71}$$

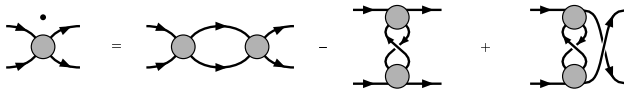


Figure 2: The RGDE for f

4.2 Spin rotation invariance

We now derive the consequences of $SU(2)$–invariance (or, more generally, $SU(N)$–invariance).

The initial interaction of important many–fermion models has the $SU(2)$ spin invariance. For instance, the initial Hubbard interaction, and interactions of the form $S_x \cdot S_y$, where $S_x = \bar{\psi}(x)\frac{\sigma}{2}\psi(x)$ is the spin at x, have this property. In the Gross–Neveu model, the interaction is $U(N)$–invariant for a model with N colors.

Restricting to 1PI vertex functions that have the same invariance excludes spontaneous symmetry breaking. In general this is a further assumption. But if the model is two-dimensional, there is no true long-range order (LRO) in the system at any positive temperature by the Mermin-Wagner theorem (a version of this theorem applying to Hubbard models was proven by Koma and Tasaki [8]). An effective action that is not invariant under $SU(2)$ spin rotations would automatically lead to LRO, and hence misleading results. Thus in two dimensions, the invariance is not an additional assumption, and it is very important to keep it in the effective interaction and the 1PI functions. The Kosterlitz–Thouless–like behaviour that is expected in two–dimensional superconductors below T_c is possible with spin-invariant interactions; they simply get a slow power-law falloff (in layered materials, there is always a coupling in the third direction which then stabilizes superconductivity).

Spin rotation invariance restricts the form of f as follows. If we define a spin tensor (cf. (24))

$$F(s \mid x_1, .., x_4)_{\sigma_1..\sigma_4} = f(s \mid (x_1, \sigma_1), .., (x_4, \sigma_4)) \tag{72}$$

then

$$\begin{aligned} F(s \mid x_1, \ldots, x_4) = \; & - \quad \varphi(s \mid x_1, x_2, x_3, x_4)D \\ & + \quad \tilde{\varphi}(s \mid x_1, x_2, x_3, x_4)E \end{aligned} \tag{73}$$

where $D_{\sigma_1\ldots\sigma_4} = \delta_{\sigma_1\sigma_4}\delta_{\sigma_2\sigma_3}$ and $E_{\sigma_1\ldots\sigma_4} = \delta_{\sigma_1\sigma_3}\delta_{\sigma_2\sigma_4}$. This can be seen by the following argument (which also applies to $U(N)$–symmetries. The symmetry transforms the fields as $\psi \to U\psi$, $\bar\psi \to \bar U\bar\psi$. Considering an infinitesimal transformation, we get see that the only invariants are given by the above Kronecker deltas E and D (also in the $U(N)$ case).

The equation $E_{\sigma_2\sigma_1\sigma_3\sigma_4} = D_{\sigma_1\sigma_2\sigma_3\sigma_4}$ and the antisymmetry of f under $(x_1,\sigma_1) \leftrightarrow (x_2,\sigma_2)$ imply that

$$\begin{aligned}\tilde\varphi(s \mid x_1,x_2,x_3,x_4) &= \varphi(s \mid x_2,x_1,x_3,x_4)\\ &= \varphi(s \mid x_1,x_2,x_4,x_3)\end{aligned} \tag{74}$$

Exchanging twice, we have (similarly to (74))

$$\varphi(s \mid x_2,x_1,x_4,x_3) = \varphi(s \mid x_1,x_2,x_3,x_4)\,. \tag{75}$$

However there is no symmetry of φ under exchange of only one pair of coordinates.

That interactions of the form $S_x \cdot S_y$ can be written in the form (73) follows from the Fierz identity

$$\sum_{i=1}^{3}(\sigma^i)_{\mu\nu}(\sigma^i)_{\alpha\beta} = 2\delta_{\alpha\nu}\delta_{\beta\mu} - \delta_{\mu\nu}\delta_{\alpha\beta}. \tag{76}$$

Using (76), one can also reconstruct the four–fermion interaction in the form $\tilde S\cdot\tilde S + \tilde\rho\tilde\rho$, where $\tilde S$ and $\tilde\rho$ transform like spin densities and charge densities as concerns the spin dependence. For a general φ, $\tilde S$ and $\tilde\rho$ will not be local in space but involve different points.

We now rewrite the RGDE in terms of φ. The spin sum in the particle-particle term is of the form

$$\sum_{\tau,\tau'} A_{\sigma_1\sigma_2\tau\tau'}B_{\tau'\tau\sigma_3\sigma_4} = (A*B)_{\sigma_1\sigma_2\sigma_3\sigma_4} \tag{77}$$

and the spin sum in the particle-hole term is of the form

$$\sum_{\tau,\tau'} A_{\sigma_1\tau\tau'\sigma_4}B_{\tau'\sigma_2\sigma_3\tau} = (A\circ B)_{\sigma_1\sigma_2\sigma_3\sigma_4} \tag{78}$$

where A and B are F with different spatial arguments. Straightforward calculation gives the relations

$$\begin{aligned}D*D &= D,\ D*E = E*D = E,\ E*E = D,\\ D\circ D &= 2D,\ D\circ E = E\circ D = D,\ E\circ E = E\,.\end{aligned} \tag{79}$$

Inserting (73) into (67), using (79), and comparing the coefficient of $-D$, we get the RGDE in the form

$$\dot{\varphi}(s) = \mathcal{T}_{\rm pp}(s) + \mathcal{T}^{\rm d}_{\rm ph}(s) + \mathcal{T}^{\rm cr}_{\rm ph}(s) \tag{80}$$

where, using $\underline{x} = (x_1, x_2, x_3, x_4)$,

$$\begin{aligned} \mathcal{T}_{\rm pp}(s|\underline{x}) = -\int {\rm d}y_1 \ldots {\rm d}y_4 L(y_2, y_1, y_3, y_4) \\ \varphi(s|x_1, x_2, y_2, y_3)\varphi(s|y_4, y_1, x_3, x_4), \end{aligned} \tag{81}$$

$$\begin{aligned} \mathcal{T}^{\rm d}_{\rm ph}(s|\underline{x}) = -\int {\rm d}y_1 \ldots {\rm d}y_4 L(y_1, y_2, y_3, y_4) \\ [\,-2\varphi(s|y_2, x_2, x_3, y_3)\,\varphi(s|x_1, y_4, y_1, x_4) \\ +\,\varphi(s|y_2, x_2, x_3, y_3)\,\varphi(s|y_4, x_1, y_1, x_4) \\ +\,\varphi(s|x_2, y_2, x_3, y_3)\,\varphi(s|x_1, y_4, y_1, x_4)], \end{aligned} \tag{82}$$

$$\begin{aligned} \mathcal{T}^{\rm cr}_{\rm ph}(s|\underline{x}) = -\int {\rm d}y_1 \ldots {\rm d}y_4 L(y_1, y_2, y_3, y_4) \\ \varphi(s|x_2, y_2, x_4, y_3)\,\varphi(s|y_4, x_1, y_1, x_3), \end{aligned} \tag{83}$$

and $L(y_1, y_2, y_3, y_4) = S(y_1, y_2)G(y_3, y_4) + G(y_1, y_2)\; S(y_3, y_4)$. Similarly, the equation for the full inverse two–point function is

$$\dot{\gamma}_2(s\; x_1, x_2) = \dot{q}_s(x_1, x_2) - \dot{\Sigma}_s(x_1, x_2) \tag{84}$$

with a scale–dependent selfenergy Σ_s that satisfies

$$\begin{aligned} \dot{\Sigma}_s(x_1, x_2) \;\; = \;\; & \int {\rm d}x_3\, {\rm d}x_4 S_s(x_4, x_3) \\ & [\varphi(s|x_3, x_1, x_4, x_2) - 2\varphi(s|x_1, x_3, x_4, x_2)] \end{aligned} \tag{85}$$

The initial condition for Σ_s depends on how the Fermi surface is renormalized.

Again, these equations have a convenient graphical interpretation: φ was the coefficient of the direct spin term D, where the spin of particle 1 is the same as that of particle 4, and similarly for 2 and 3. Thus we may draw the vertex as in Figure 3, where the solid fermion lines going through at the top and the bottom of the vertex indicate that spin is conserved along

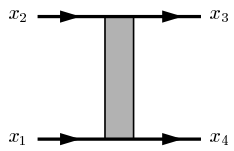


Figure 3: The vertex corresponding to $\varphi(s|x_1, x_2, x_3, x_4)$

these lines. The symmetry (75) means that it does not matter if the vertex is drawn upside down. The contributions to the right hand side of (80) can then be represented graphically as in Figure 4. Thus (80) and (84) are obtained by the usual QED–like Feynman rules, and the factor 2 and the minus sign appear as the usual factors from the spin trace and the extra minus sign of the fermion loop.

Note, however, that this graphical correspondence does not mean that the fermion–boson vertex function that one can associate to the interaction by a Hubbard–Stratonovitch transformation is independent of momentum, as it would be for the bare Hubbard interaction. On the contrary: its momentum dependence will be very important. Also, the propagators associated to the internal lines of these graphs are the full propagator $G(x, y)$ and the single–scale propagator $S(x, y)$. The function L is the symmetric combination of the two, which means that every diagram stands for the two contributions given by the summands in L.

The symmetry (10) implies that

$$\varphi(s|x_1, x_2, x_3, x_4) = \overline{\varphi(s|Rx_4, Rx_3, Rx_2, Rx_1)} \tag{86}$$

where $R(\tau, \mathbf{x}) = (-\tau, \mathbf{x})$. Similarly, the selfenergy satisfies

$$\Sigma_s(x_1, x_2) = \overline{\Sigma_s(Rx_1, Rx_2)}. \tag{87}$$

The contributions to the right hand side of (85) have the graphical representation shown in Figure 5, where the internal line stands for a single scale propagator.

4.3 Translation invariance

If translation invariance is unbroken, we can take the Fourier transform. In contrast to charge and spin invariance, translation invariance is only discrete

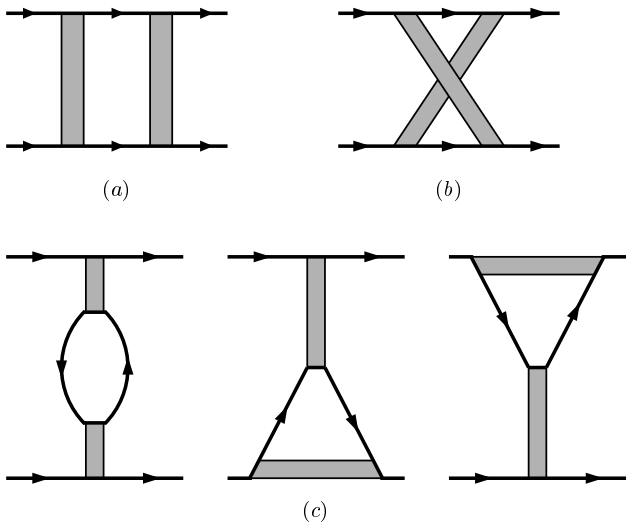


Figure 4: The contributions to the right–hand side of the RGDE. (a) the particle–particle term (b) the crossed particle–hole term (c) the direct particle–hole term; the first of the three graphs gets a factor -2 because of the fermion loop.

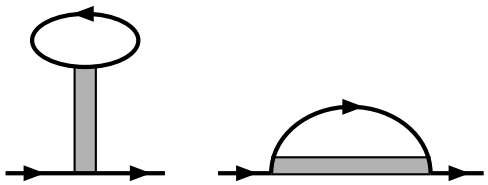


Figure 5: The contributions to the selfenergy

in our lattice model and thus may be broken at positive temperature also in two dimensions. Thus specializing to unbroken translation invariance is a further assumption. It can be relaxed if one assumes that invariance under a sufficiently large subgroup, e.g. that of translations of a sublattice, still holds. The corresponding Fourier transform is then defined on a smaller momentum space.

We take the convention that momenta corresponding to $\bar{\psi}$ are counted outgoing and those corresponding to ψ are counted as incoming. Then translation invariance implies that $\hat{\varphi}(s|p_1, p_2, p_3, p_4) = \delta(p_1 + p_2, p_3 + p_4)\, V_s(p_1, p_2, p_3)$, and the equation for V_s reads

$$\dot{V}_s = \hat{\mathcal{T}}_{\rm pp} + \hat{\mathcal{T}}^{\rm d}_{\rm ph} + \hat{\mathcal{T}}^{\rm cr}_{\rm ph}, \tag{88}$$

with the particle–particle term

$$\begin{aligned} \hat{\mathcal{T}}_{\rm pp}(p_1, p_2, p_3) = -\int \mathrm{d}k \mathcal{L}_-(p_1 + p_2, k) \\ V_s(p_1, p_2, k) V_s(p_1 + p_2 - k, k, p_3), \end{aligned} \tag{89}$$

the direct particle–hole term

$$\begin{aligned} \hat{\mathcal{T}}^{\rm d}_{\rm ph}(p_1, p_2, p_3) = -\int \mathrm{d}k\ \mathcal{L}_+(p_2 - p_3, k) \\ [\, -2V_s(k, p_2, p_3)\, V_s(p_1, p_2 - p_3 + k, k) \\ +V_s(k, p_2, p_3)\, V_s(p_2 - p_3 + k, p_1, k) \\ +V_s(p_2, k, p_3)\, V_s(p_1, p_2 - p_3 + k, k)], \end{aligned} \tag{90}$$

and the crossed particle–hole term

$$\begin{aligned} \hat{\mathcal{T}}^{\rm cr}_{\rm ph}(p_1, p_2, p_3) = -\int \mathrm{d}k\ \mathcal{L}_+(p_3 - p_1, k) \\ V_s(p_2, k, p_1 + p_2 - p_3)\, V_s(p_3 - p_1 + k, p_1, k). \end{aligned} \tag{91}$$

Here

$$\mathcal{L}_\pm(q, k) = \hat{S}(k)\hat{G}(q \pm k) + \hat{S}(q \pm k)\hat{G}(k). \tag{92}$$

The symmetry (86) implies that

$$V_s(p_1, p_2, p_3) = \overline{V_s(R(p_1 + p_2 - p_3), Rp_3, Rp_2)} \tag{93}$$

with $R(\omega, \mathbf{p}) = (-\omega, \mathbf{p})$.

These RG equations were studied in [9], and similar ones in [5] and [4]. For the numerical solution, there are still too many variables. By projecting on the Fermi surface one can reduce the number of variables in the equations; to calculate response functions one then calculates the flow of susceptibility vertices, which is driven by that of the coupling functions. Details about the susceptibilities are given in the appendix.

4.4 Comparison to other RG schemes

The main differences of this approach to the one in [4, 5] are the following. In [4], Polchinski's equation was used. The quantities appearing in that equation are the connected amputated Green functions, which are in general 1–particle reducible. Therefore, Zanchi and Schulz[4] have to keep the tree–level six–point function and insert it into the equation for the four–point function, thereby obtaining a flow equation that is nonlocal in the flow parameter ϵ_s. This is the normal procedure in integrating the flow equation and, when iterated out, it gives a very useful representation for rigorous studies, pioneered in [18]. We note that our truncation of the RGDE system, namely dropping the *one–particle irreducible* six–point function from the flow, actually also retains the one–particle reducible six–point function at tree–level. Thus in that respect, the two formalisms are similar. However, in the 1PI formalism, the higher n–point functions are also all kept at tree level automatically, and the flow equation is local.

In [5], Halboth and Metzner use the Wick ordered RGDE of [2], which has a number of advantages. First, as in the 1PI scheme, the flow equation is local; the influence of the six–point function on the four–point function is taken into account automatically by Wick ordering. The Wick ordering also implies that part of the lines are supported *below* the scale ϵ_s instead of above it. This makes the justification of projections to the Fermi surface easier and it also leads to rather clear thresholds where scattering processes die out because the available phase space disappears. One disadvantage of the Wick ordered formalism is that taking selfenergy corrections into account is not as convenient as in the 1PI formalism. If self–energy corrections are ignored, as has been done in [5] and [4], and mostly in [9], this disadvantage disappears. However, in the presence of van Hove singularities, self–energy corrections may be important. So far, they have only been included in the calculation of the Fermi surface deformation reported in the appendix of [9].

5 Geometric arguments for dropping the 1PI six–point function

In this section, we discuss the effects of the geometry of the energy shells which form the support of the single scale propagators on power counting. For the case of a smooth and curved Fermi surface, we use phase space bounds to estimate the contribution of the 1PI six–point function γ_6 to $\dot{\gamma}_4$. We show that in a specific scale range, the correction due to γ_6 is small even if the scale–dependent coupling constant is not small any more. We then discuss the case of Fermi surfaces with van Hove singularities.

5.1 The term of order $\gamma_4(s)^3$

We now focus on Fermi surfaces that are smooth and curved. We give this discussion for the flow with a fixed Fermi surface; the Fermi surface deformation and the field strength renormalization can be dealt with using the methods of [11, 12] and therefore not essential in this argument.

To estimate the influence of γ_6 on γ_4, we look at the integrated version of the equation (53) for $\gamma^{(6)}$. Eq. (53) has the graphical representation shown in Figure 6, where the m–legged vertex labelled by s represents $\gamma^{(m)}(s)$, and where we have also labelled the propagators associated to the lines. When coefficients of the monomials in the fields are compared to get the corresponding equations for $\gamma_6(s|X_1,\ldots,X_6)$, the antisymmetrization condition implies that there is a sum over different possibilites of associating the internal propagators to the different lines, as in (60).

The vertex functions $\hat{\gamma}_m(s|p_1,\ldots,p_{m-1})$ have a natural power counting that is determined by the scaled propagator, or equivalently, the kinetic energy. To be specific, we choose the scaled propagator of the form

$$C_s(p) = \frac{1}{\mathrm{i}\omega - e(\mathbf{p})}\,\chi\left(\frac{\omega^2 + e(\mathbf{p})^2}{{\epsilon_s}^2}\right), \tag{94}$$

where $\epsilon_s = \epsilon_0 \mathrm{e}^{-s}$, with ϵ_0 some fixed energy scale, and where χ is a smooth and monotonically increasing cutoff function with $\chi(x) = 1$ for $x \geq 1$ and $\chi(x) = 0$ for $x \leq 1/4$. The details of the choice of χ play no role. The discussion of flows in which the cutoff function χ is frequency–independent

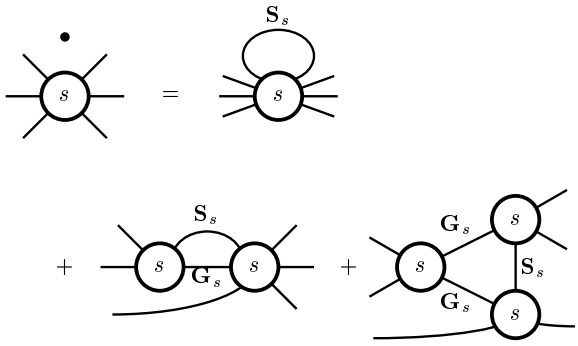


Figure 6: The differential equation for $\gamma^{(6)}$

(i.e. where $\chi(e(\mathbf{p})^2/{\epsilon_s}^2)$ is the cutoff function) is similar – the power counting is essentially the same.

The essential properties of $\dot{C}_s$ are that it is large near to the Fermi surface, but zero outside of a thin shell in $(d+1)$–dimensional momentum space around the Fermi surface. More precisely, because the differentiated cutoff function $\chi'(x)$ is nonzero only for $1/4 \le x \le 1$, $\dot{C}_s(k)$ vanishes unless $\epsilon_s/2 \le |\mathrm{i}\omega - e(\mathbf{k})| \le \epsilon_s$, so

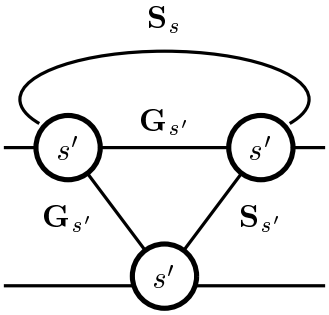
$$|\dot{C}_s(k)| \le \frac{2}{\epsilon_s} 1_s(k) \tag{95}$$

where 1_s is the indicator function $1_s(k) = 1$ if k is in the energy shell where $\dot{C}_s \neq 0$ (i.e. if $\epsilon_s/2 \le |\mathrm{i}\omega - e(\mathbf{k})| \le \epsilon_s$), and $1_s(k) = 0$ if k is not in that shell. In terms of the density of states $N(E) = \int \frac{\mathrm{d}^d\mathbf{k}}{(2\pi)^d}\, \delta(E - e(\mathbf{k}))$, the volume of this momentum space shell is

$$\begin{aligned} W(s) &= \int \mathrm{d}^{d+1}k\, 1_s(k) \le \frac{1}{\beta} \sum_{|\omega| \le \epsilon_s} \int_{-\epsilon_s}^{\epsilon_s} \mathrm{d}E\, N(E) \\ &\le A{\epsilon_s}^2, \end{aligned} \tag{96}$$

with the constant $A = \frac{2}{\pi} N_{\max}$, where $N_{\max}$ is the maximum of the density of states $N(E)$ over all $|E| \le \epsilon_s$. Because $N(E)$ is a surface integral over $1/|\nabla e|$, the constant A is finite only in absence of a van Hove singularity.

Integrate (53) from scale 0 to s. This gives, among others, the third order correction shown in Figure 7 to $\dot{\gamma}^{(4)}(s)$. We consider this two–loop integral



(a)

Figure 7: A term cubic in γ_4 arising from the contribution of γ_6 to the flow of γ_4. This graph is integrated over s' from $s' = 0$ to $s' = s$.

first and come back to the full equation for γ_6 below. Let $\gamma_4(s) = V_s$ and let

$$g(s) = \sup_{0 \le s' \le s} \sup_{p_1,p_2,p_3} |\gamma_4(s'|p_1, p_2, p_3)| \tag{97}$$

be the maximal possible value of the four–point vertex at any flow time up to s (i.e. at energy scales $\epsilon_{s'} \ge \epsilon_s$), and for any values of the external momenta. The quantity $g(s)$ provides a bound for the largest possible four–point coupling constant that can arise in the flow. Because $g(s)$ is the maximum over $s' \le s$, it is increasing in s. Because $g(s)$ includes the maximum over all momenta, we can bound all three vertex factors $\gamma_4(s')$ by $g(s')$. The contribution I_a of the graph in Figure 7 is at most

$$I_a \le \int_0^s \mathrm{d}s' g(s')^3 \, J(s, s', q_1, q_2) \tag{98}$$

with

$$\begin{aligned} J(s, s', q_1, q_2) &= \int \mathrm{d}k \int \mathrm{d}p \, |\mathbf{S}_{s'}(k)\mathbf{S}_s(p)| \\ &| \quad \mathbf{G}_{s'}(\pm k \pm p + q_1)\mathbf{G}_{s'}(\pm k + q_2)| \end{aligned} \tag{99}$$

where q_1 and q_2 are linear combinations of the external momenta. They will be unimportant because our estimate for J will be independent of q_1 and q_2. Because $g(s)$ is an increasing function of s, we can take it out of the integral to get

$$I_a \le g(s)^3 \int_0^s \mathrm{d}s' \, J(s, s', q_1, q_2) \tag{100}$$

The contribution is third order in $g(s)$; if $g(s)$ gets large, it will only stay smaller than the second order terms if the coefficient function J is small. Write

$$\mathbf{G}_{s'}(\pm k \pm p + q_1) = \int_0^{s'} \mathrm{d}s'' \dot{\mathbf{G}}_{s''}(\pm k \pm p + q_1). \tag{101}$$

Because $\dot{\mathbf{S}}_s$ and $\dot{\mathbf{G}}_s$ have the same size and support properties as $\dot{C}_s$, we can use (95) to get

$$J(s, s', q_1, q_2) \leq \frac{1}{\epsilon_s {\epsilon_{s'}}^2} \int \frac{\mathrm{d}s''}{\epsilon_{s''}} \ \mathrm{Vol}(s, s', s'') \tag{102}$$

with the two–loop momentum space volume

$$\mathrm{Vol}(s, s', s'') = \int \mathrm{d}^{d+1}k \quad \int \quad \mathrm{d}^{d+1}p \, 1_s(p) \, 1_{s'}(k) \\ 1_{s''}(\pm k \pm p + q_1) \tag{103}$$

which depends on the three scales s, s', s'' and on q_1. Here we have also used that $1_{s'}(\pm k + q_2) \leq 1$, so the dependence on q_2 has already dropped out. One can easily bound J by dropping the third indicator function from the integral. This gives

$$\mathrm{Vol}(s, s', s'') \leq W(s) W(s'), \tag{104}$$

with $W(s)$ as in (96), so

$$J(s, s', q_1, q_2) \leq A^2 \epsilon_s \int_0^{s'} \frac{\mathrm{d}s''}{\epsilon_{s''}} \tag{105}$$

Because $\epsilon_s = \epsilon_0 \mathrm{e}^{-s}$, the last integral is bounded by $1/\epsilon_{s'}$. Inserting this back into (100) and doing the integral over s', we arrive at the estimate

$$I_a \leq A^2 g(s)^3, \tag{106}$$

which does not suggest that this term suppressed. This would correspond to usual power counting, which suggests that all four–legged contributions have the same order of magnitude.

However, if one does not drop $1_{s''}(\pm k \pm p + q_1)$ from the above integral, one gets a better bound. It was shown in [11, 12] (see also Appendix B.8 of [3]) that

$$\mathrm{Vol}(s, s', s'') \leq W(s) W(s') Q_{\mathrm{vol}} (1 + s) \epsilon_{s''} \tag{107}$$

where $Q_{\rm vol}$ is a constant that depends on the Fermi surface geometry and which we shall discuss below. This bound does not depend on q_1. In $d \geq 3$, the logarithmic factor $1+s$ is replaced by 1. Compared to (104), there is an extra factor $Q_{\rm vol}(1+s)\epsilon_{s''}$ in (107), and this factor is small for s'' large. Integrating the scales in the same way as before, we now get

$$I_a \leq A^2 Q_{\rm vol} g(s)^3 \epsilon_s (1+s)^2 \tag{108}$$

which, up to the factor $(1+s)^2$, is $\epsilon_s \sim {\rm e}^{-s}$ smaller than before.

Let us now calculate the correction that this gives. In the flow where the influence of γ_6 is omitted, we have

$$\dot{g}(s) \leq \tilde{\beta}_2 g(s)^2 \tag{109}$$

with $\tilde{\beta}_2 > 0$ the maximal value of the single–scale bubble integral. Here we have taken bounds; this gives the worst case scenario; if one looks at the flow of a particular coupling constant (in an expansion in angular momentum couplings, e.g. the s–wave coupling), the sign of the coupling constant at $s=0$ determines whether it grows or decreases under the flow. Taking the equality sign in (109), we get

$$g(s) = \frac{g(0)}{1 - g(0)\tilde{\beta}_2 s}, \tag{110}$$

which diverges at some nonzero s_c, indicating the possibility of instabilities. The flow given by (110) has essentially two different regimes. If $g(0)\tilde{\beta}_2 \ll 1$, then $g(s) \sim g(0) + g(0)\tilde{\beta}_2 s$ grows logarithmically in ϵ_s. When s approaches s_c, it starts to grow like the inverse power $1/(s_c - s)$.

At a positive temperature $T = 1/\beta$, the flow stops at $s_\beta = \log \frac{\beta \epsilon_0}{\pi}$, because nothing is left to integrate over when $\epsilon_s = \pi/\beta$. If $g(o)$ is so small that $s_c > s_\beta$, the four–point function stays finite, and there is no instability. This is the basis of the Fermi liquid criterion of [2]: if $s_\beta \ll s_c$, the running coupling constants stay small and one can use convergent perturbation theory to justify Fermi liquid theory.

But even if the coupling constants do not stay small, the correction I_a to the $O(g(s)^2)$ four–point flow is small if

$$A^2 Q_{\rm vol} g(s) \epsilon_s (1+s)^2 \ll 1 \tag{111}$$

with $g(s)$ given by (110).

Thus the flow is naturally split into three different regimes:

(i) high scales s is so small that $A^2 Q_{\rm vol} \epsilon_s (1+s)^2$ is not very small.

(ii) intermediate scales s is such that (111) holds

(iii) low scales s is so large that (111) fails.

This provides a criterion when the flow is accurate and when it has to be stopped, and it also makes clear what the role of weak coupling is in our treatment. A weak initial coupling constant is needed to get through regime (i) with a second order flow because the only small factor in region (i) is the coupling $g(s)$. If the coupling constant is not weak, there can be significant corrections to the flow in that regime. In regime (ii), the coupling constant need not be small because the scaling provides small factors that make the correction term small. In regime (iii) the corrections become dominant, so the one–loop flow has to be stopped before one enters regime (iii). What one expects to happen around this scale is that gaps open up that cut off the singularity.

If the initial interaction is sufficiently weak, more precisely, if $g(0) \log \epsilon_0$ is small enough, regimes (i) and (ii) will exist. This is because regime (i) is, by definition, the regime with infrared cutoff ϵ_0, and it has been proven [13] that perturbation theory for the many–fermion system converges if there is an infrared cutoff. Thus, if the coupling constant is small enough, low order perturbation theory can be used to calculate the effective interaction. After that, the truncated flow can be used in regime (ii). This flow decouples into angular momentum sectors (or their generalization in absence of spherical symmetry; see e.g. [10] or [3]) which flow independently. Ones that start out positive will decrease to smaller positive values, ones that start out negative will decrease to more negative values, and eventually, at some $\tilde{s}_c$, diverge in the one–loop flow. As discussed, there is no real divergence of a coupling constant in the full flow; instead, one enters regime (iii) where the corrections to the one–loop flow can no longer be neglected.

It might happen that in the flow, all coupling constants remain positive, which would mean that none of them diverges. We do not know of any example for such behaviour in $d \geq 2$ and with $\mathbf{p} \to -\mathbf{p}$ symmetry – there is always a coupling constant that becomes negative (Kohn–Luttinger effect) at scale ϵ_0 and therefore starts to grow in absolute value at lower scales. If the symmetry under $\mathbf{p} \to -\mathbf{p}$ is broken, there is no Cooper instability and the coupling constants remain finite down to scale zero even at zero temperature

[15]. The reason for this is not that all couplings stay positive (in fact, their sign does not matter), but that the constant $\tilde{\beta}_2$ is replaced essentially by const. ${\epsilon_s}^{1/3}$ (see Lemma 4.8 in [12]), which just integrates to a constant, so that $g(s) \leq g(0)/(1 - g(0)$ const.) in that case.

5.2 The influence of the full γ_6

The above discussion was only for a single graph, but the essential points also apply to the other terms, because of the following. The equation for γ_6 is linear; in finite volume, where the momenta are discrete, it can be viewed as a system of linear differential equations, labelled by $\underline{P} = (P_1, \ldots, P_6)$, where the P_i include momentum, spin, and charge indices. We have

$$\begin{aligned} \dot{\gamma}_6(s) &= B_6(\gamma_4(s), \gamma_6(s)) + T_6(\gamma_4(s), \gamma_4(s), \gamma_4(s)) \\ &+ L_6(\gamma_8(s)) \end{aligned} \tag{112}$$

where B_6, T_6, and L_6 are linear in each of the coupling functions, and depend on the momenta $\underline{P}$. We drop γ_8 from the equation. The argument for doing this is similar to the one we are describing (see below). Now γ_6 can be calculated in terms of γ_4 as the solution of the linear differential equation (112). Defining a maximal six–point coupling $g_6(s)$ (independent of momentum) in analogy to (97), one gets

$$\dot{g}_6(s) \leq \tilde{\beta}_2 g(s) g_6(s) + \frac{1}{\epsilon_s} g(s)^3. \tag{113}$$

To get a bound for g_6, we replace the inequality by an equality. This is now just a linear differential equation in one variable s, which can be solved explicitly. Using that $g(s)$ is increasing in s, we get

$$g_6(s) \leq \frac{1}{\epsilon_s} g(s)^3 \exp\left(\tilde{\beta}_2 \int_0^s g(s')\, \mathrm{d}s'\right). \tag{114}$$

Here we already see the main change as compared to the discussion of the third order term: the integral of $g(s)$ gets exponentiated. Using (110), we can calculate this integral and get

$$g_6(s) \leq \frac{1}{\epsilon_s} g(s)^3 \, (1 - g(0)\tilde{\beta}_2 s)^{-1}. \tag{115}$$

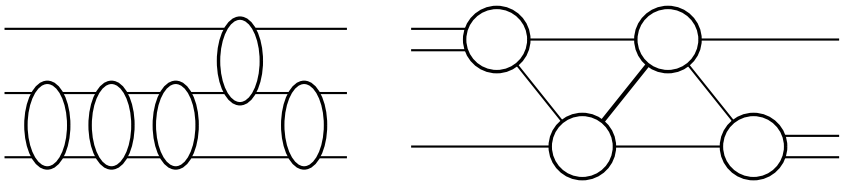


Figure 8: Graphical representation of terms arising in the solution of the differential equation for γ_6 as a function of γ_4. (a) a chain of four-legged insertions attached to the third–order term in the equation for γ_6 (b) a term containing overlapping loops which has a small scaling factor

In other words, solving the differential equation for γ_6 gives another factor in the denominator which is similar to the one already in $g(s)$. Thus the power in the denominator increases by one.

In graphical terms, the integration of the differential equation for γ_6 corresponds to a resummation of infinitely many graphs, and the new denominator comes from long chains of four–point functions attached to the third order term, as shown in Figure 8 (a), which arise from iterating the same term in the equation many times.

The other terms, such as the one drawn in Figure 8 (b), contain a loop overlap [11] on one line (e.g. the one drawn heavy in the figure) and are therefore suppressed (this is also the reason why the prefactor of $g(s)g_6(s)$ in (113) is $\tilde{\beta}_2$). If one goes through this in more detail (which we skip here for brevity, and because the argument is similar to the one in [11]), one finds that the overlapping loop structure is in all terms that do not involve self–contractions of four–point vertices. The latter are treated by a Fermi surface counterterm in the way done in [11, 12].

The backreaction of γ_6 on γ_4 is now obtained by joining two legs of different vertices to form an additional loop. The above discussion has shown that this always creates a loop overlap which cancels the $1/\epsilon_s$ in (115). Hence the correction of γ_6 to the flow for γ_4 is bounded by

$$\psi(s, g(s)) = Q_{\text{vol}} A^2 (1+s)^2 g(s)^3 \frac{1}{1 - g(0)\tilde{\beta}_2 s}. \tag{116}$$

Thus, the analogue of (111), with the power of the denominator increased by

one, defines the condition that the correction remains small.

The above argument has shown that, with $g(s)$ given by the unperturbed one–loop flow (110), the term $\psi(s, g(s))$ that we dropped is small in regions (i) and (ii). A standard stability analysis now shows that the solution of the equation where the contribution of γ_6 is taken into account in the equation for γ_4, is given by $g(s) \to g(s)(1 + h(s))$, where $h(s)$ is small as long as $\psi(s, g(s))$ remains small.

This completes the justification of dropping γ_6 in the equation for γ_4 under the hypothesis that γ_8 can be dropped from the equation for γ_6. One can now go on to discuss the equations for γ_8 etc. The same argument applies because of the following observations: (1) when γ_{m+2} is dropped, γ_m can be expressed in terms of γ_4 by integration of the differential equation. (2) the γ_m are 1PI functions, so the selfcontraction term which γ_m contributes to γ_{m-2} always involves a two–loop graph which is either taken care of by Fermi surface renormalization or gets a small factor by the overlapping loop bound. A careful graphical analysis is done in [11, 12]. The combinatorial problem is treated in [13].

5.3 Fermi surfaces with van Hove singularities

We now discuss what changes if van Hove singularities and small curvature of the Fermi surface are present. Obviously, the estimate (96) for $W(s)$ no longer holds if the region of $\mathbf{k}$ with $|e(\mathbf{k})| \leq \epsilon_s$ contains a zero of the gradient. In $d = 2$, the integral for $\mathbf{k}$ at distance at most k_c from a saddle point of e is, in natural momentum space coordinates (u, v) around the saddle point, where $e(\mathbf{p}) = uv$,

$$\int\limits_{|e(\mathbf{k})| \leq \epsilon_s} \frac{\mathrm{d}^d \mathbf{k}}{(2\pi)^d} \sim \int\limits_{\substack{|uv| \leq \epsilon_s \\ u^2+v^2 \leq k_c^2}} \mathrm{d}u \; \mathrm{d}v \sim \epsilon_s \log \frac{k_c^2}{\epsilon_s} \tag{117}$$

so there is an extra logarithmic factor s. Again, for $d \geq 3$ this factor is replaced by 1. Moreover, Q_{vol} depends on the curvature of the Fermi surface and diverges if the curvature vanishes.

For our case this means that the argument for omitting higher order terms works only on those parts of the Fermi surface which are curved (and that regime (i) can get uncomfortably large). That the curvature is nonzero

is meant in the more precise sense that the scale should be so low that $Q_{\rm vol}(1+s)\epsilon_s < 1$. Otherwise, (107) provides no improvement over (104). In the $t' = 0$ Hubbard model near to half–filling (μ small), the improvement starts only at scales $\epsilon_s < \mu$. For the (t, t') model with $t' \neq 0$, the curvature is nonvanishing away from the saddle points. Moreover, the curved region of the Fermi surface gets larger and larger as the scale decreases because the shells get thinner. Thus one can take the radius k_c of the disk around the van Hove singularity to shrink with ϵ_s, and the part of momentum space where there is no improvement gets smaller and smaller. Inside the disk of radius k_c, there is, however, no gain. Choosing k_c proportional to $\sqrt{\epsilon_s}$, we get that the influence of a region of that size to the integral is of order 1 instead of order s. Thus the origin of the extra logarithms is not in a very small neighbourhood of the van Hove singularity but in the range of $\sqrt{\epsilon_s} \leq k_c \leq 1$. This is similar to the observation that the one–dimensional integral $\int_0^1 \frac{{\rm d}x}{\sqrt{x^2+a^2}}$, which grows like $|\log a|$ for small a, can be split into the contribution of a small neighbourhood of 0, $\int_0^a \frac{{\rm d}x}{\sqrt{x^2+a^2}}$, which is bounded because it is $\leq \int_0^a \frac{{\rm d}x}{a} \leq 1$ and the piece $\int_a^1 \frac{{\rm d}x}{\sqrt{x^2+a^2}} = \int_a^1 \frac{{\rm d}x}{x} +$ regular terms, which produces the logartihm.

In presence of a van Hove singularity, the estimate for the leading flow, i.e. the largest coupling changes to

$$g(s) = \frac{g(0)}{1 - g(0)\tilde{\beta} s^2} \tag{118}$$

because of the extra logarithm from the local density of states integral (117), which effectively replaces $\tilde{\beta}_2$ by $\tilde{\beta}_2 s$ in the differential equation. The flow now diverges unless $T \geq T_0 {\rm e}^{-1/\sqrt{g(0)\tilde{\beta}_2}}$, which is much earlier than in absence of a van Hove singularity. It should be kept in mind, however, that in the weakly coupled repulsive Hubbard model, any attractive coupling generated by integrating out degrees of freedom is of order at least U^2, and therefore the square root just gives back the usual scaling of T as a function of U.

A Projections and susceptibilities

Even the truncated system of RG equations cannot be solved exactly in the physically interesting cases. The four–point function is a function of three

momenta and three frequencies (the fourth being fixed by momentum conservation) thus a function of $3d+3$ variables, and the RGDE is a nonlinear integrodifferential equation for this function. Similarly, the selfenergy depends on momentum and frequency. The large number of variables makes a computational solution of the RGDE difficult. However, the RG and its power counting provide further means for reducing the number of variables. The basic argument is, again, that the singularity is on the Fermi surface, and therefore only the degrees of freedom on the Fermi surface should influence the leading behaviour. A more detailed argument uses Taylor expansions around the Fermi surface. Although often discussed as a standard RG argument, a rigorous justification of this procedure is nontrivial, in particular in the case with van Hove singularities.

The projected four–point function does not contain the full information any more, and in particular it is not sufficient for calculating response functions to determine instabilities. These response functions can be recovered by considering a flow with external fields and calculating the linear and quadratic term in the external field.

The general feature of these equations is that only the equation for the coupling function (i.e. the projected function) is nonlinear, and that, given the coupling function, the others can be obtained from a linear equation whose coefficients depend on the coupling function. In this way, the logarithmic growth of couplings integrates to power laws of the response functions, as well as for the four–point function away from the Fermi surface.

The susceptibilities are obtained by coupling external boson fields a to the bilinears in the fermions that represent charge, spin, Cooper pair and other local densities, and by calculating the corresponding RG flow for these functions. The presence of the a fields makes no difference for the derivation of the RGDE because the a fields simply appear as parameters. Thus the equations are unchanged, except that now all functions depend on a. The fields a are external because no integration in a is done. Correlation functions for the gaps, spins, etc, are given by derivatives of the generating functional with respect to the a fields. Thus for their calculation, it suffices to get the dependence on a in terms of an expansion in a. The expansion of Γ_s now reads

$$\Gamma(s \mid a, \phi) = \sum_{m,n \geq 0} \gamma^{(m,n)}(s \mid a, \phi) \tag{119}$$

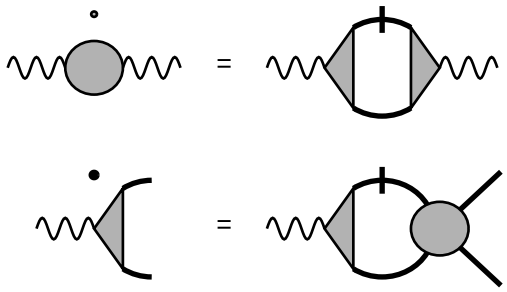


Figure 9: The differential equations for γ_{20} and γ_{12}.

with

$$\gamma^{(m,n)}(s \mid a, \phi) = \frac{1}{m!n!} \int \mathrm{d}^m \underline{X}\, \mathrm{d}^n \underline{Y} \gamma_{m,n}(s|\underline{X}, \underline{Y}) a^m(\underline{X}) \phi^n(\underline{Y})\,. \tag{120}$$

Because the a fields are boson fields, the coefficient $g_{m,n}(s|\underline{X}, \underline{Y})$ is totally symmetric under permutations of $\underline{X}$. The RGDE is now derived in the same way as above. Because the a fields are external fields only, the equations for the a–independent parts γ_{0n} remain unchanged, so that $\gamma_{0n} = \gamma_n$ for all n, with the γ_n given as above. Thus the flow for the susceptibilities is driven by the flow for the coupling functions; it takes the form of a system of linear integro–differential equations.

For $m + n > 0$ we have

$$\begin{aligned} \dot{\gamma}^{(m,n)}(s \mid a, \phi) &= -\frac{1}{2}\, \mathrm{Tr}\, [G_2 \dot{\mathbf{Q}}_s G_2 \tilde{\gamma}^{(m,n)}(s \mid a, \phi)] \\ &+ \frac{1}{2} \sum_{p \geq 2} (-1)^p \sum_{\substack{m_1, \ldots, m_p \geq 0 \\ m_1 + \ldots + m_p = m}} \sum_{\substack{n_1, \ldots, n_p \geq 2 \\ n_1 + \ldots + n_p = n}} \\ &\qquad \mathrm{Tr}\, \left[G_2 \dot{\mathbf{Q}}_s \prod_{q=1}^{p} G_2 \tilde{\gamma}^{(m_q, n_q)}(s \mid a, \phi) \right] \end{aligned} \tag{121}$$

with $\tilde{\gamma}^{(m,n)}$ defined in analogy to (44) and (45).

Since one a field couples to a fermionic bilinear, an external a field corresponds to a pair of fermions. Thus the truncation consistent with dropping

the 1PI six–point function is to leave out all m and n with $2m+n \geq 6$. This gives the equations

$$\begin{aligned} \dot{\gamma}_{12}(s|X;Y_1,Y_2) &= \frac{1}{2}\int \mathrm{d}^4\underline{Z}\,\mathbf{L}(Z_1,\ldots,Z_4) \\ &\quad \gamma_{12}(s|X,Z_2,Z_3)\gamma_{04}(s|Z_4,Z_1;Y_1,Y_2) \end{aligned} \tag{122}$$

and

$$\begin{aligned} \dot{\gamma}_{20}(s|X_1,X_2) &= \frac{1}{2}\int \mathrm{d}^4\underline{Z}\,\mathbf{L}(Z_1,\ldots,Z_4) \\ &\quad \gamma_{12}(s|X_1,Z_2,Z_3)\gamma_{12}(s|X_2;Z_4,Z_1) \end{aligned} \tag{123}$$

with $\mathbf{L}$ given by (60). The initial condition on γ_{12} at $s=0$ determines which susceptibility is considered; in particular, it determines the symmetry of the superconducting instability in the case of the coupling to Cooper pairs.

In presence of charge invariance, we get separate equations for the particle–particle ("superconductance") and particle–hole ("normal")susceptibilities, defined as

$$\begin{aligned} &\gamma_{12}^{pp,\epsilon}(s|X,(y_1,\sigma_1),(y_2,\sigma_2)) \\ &= \gamma_{12}(s|X;(y_1,\sigma_1,\epsilon),(y_2,\sigma_2,\epsilon)) \end{aligned} \tag{124}$$

and

$$\begin{aligned} &\gamma_{12}^{ph,\epsilon}(s|X,(y_1,\sigma_1),(y_2,\sigma_2)) \\ &= \gamma_{12}(s|X;(y_1,\sigma_1,\epsilon),(y_2,\sigma_2,-\epsilon)) \end{aligned} \tag{125}$$

By the fermionic antisymmetry,

$$\gamma_{12}^{pp,-}(s|X,(y_1,\sigma_1),(y_2,\sigma_2)) = -\gamma_{12}^{pp,+}(s|X,(y_2,\sigma_2),(y_1,\sigma_1)) \tag{126}$$

and similarly for $\gamma_{12}^{ph,\pm}$, so it suffices to consider one of the "$\pm$" quantities. We now also assume spin rotation invariance; then the normal charge ($\sim \delta_{\sigma_1\sigma_2}$) and spin $\sim (\tau_3)_{\sigma_1\sigma_2}$ susceptibility (τ_3 the Pauli matrix) do not couple in the flow. The resulting equations are

$$\begin{aligned} &\dot{\gamma}_{12}^{pp,-}(s|X,(y_1,\sigma_1),(y_2,\sigma_2)) \\ &= \int \mathrm{d}u_1 \ldots \mathrm{d}u_4 L(u_1,\ldots,u_4) \\ &\quad \gamma_{12}^{pp,-}(s|X,(u_2,\sigma_1),(u_3,\sigma_2))\,\varphi(s|u_4,u_1;y_1,y_2) \end{aligned} \tag{127}$$

for the superconductance susceptibility,

$$\begin{aligned} & \dot{\gamma}_{12,\mathrm{charge}}(s|x;y_1,y_2) \\ = & \int \mathrm{d}u_1 \ldots \mathrm{d}u_4 \ \mathrm{Re}\ L(u_1,u_2,u_4,u_3) \\ & \gamma_{12,\mathrm{charge}}(s|x;u_3,u_2) \\ & [2\varphi(s|y_1,u_1,u_4,y_2) - \varphi(s|u_1,y_1,u_4,y_2)] \end{aligned} \tag{128}$$

for the charge susceptibility, and

$$\begin{aligned} & \dot{\gamma}_{12,\mathrm{spin}}(s|x;y_1,y_2) \\ = & \int \mathrm{d}u_1 \ldots \mathrm{d}u_4 \ \mathrm{Re}\ L(u_1,u_2,u_4,u_3) \\ & \gamma_{12,\mathrm{spin}}(s|x;u_1,u_4)\varphi(s|u_3,y_1,u_2,y_2) \end{aligned} \tag{129}$$

for the spin susceptibility.

References

[1] J.W. Negele and H. Orland, Quantum Many–Particle Systems, Addison-Wesley, Reading, MA, 1988.

[2] M. Salmhofer, Comm. Math. Phys. **194** (1995) 249.

[3] M. Salmhofer, Renormalization, Springer Texts and Monographs in Physics, Springer, Heidelberg, 1998.

[4] D. Zanchi and H.J. Schulz, *Phys. Rev.* **B 61**, 13609 (2000)

[5] C. Halboth and W. Metzner, *Phys. Rev.* **B 61**, 7364 (2000)

[6] J. Polchinski, *Nucl. Phys.* **B 231** (1984) 269

[7] B. Simon, The Statistical Mechanics of Lattice Gases, Volume 1, Princeton University Press, Princeton, NJ, 1993

[8] T. Koma and H. Tasaki, Phys. Rev. Lett. **68** (1992) 2348

[9] C. Honerkamp, M. Salmhofer, N. Furukawa, T.M. Rice, cond-mat/9912358, to appear in *Phys. Rev. B*

[10] J. Feldman, E. Trubowitz, *Helv. Phys. Acta* **63** (1990) 157, ibid. **64** (1991) 213

[11] J. Feldman, M. Salmhofer, E. Trubowitz, *J. Stat. Phys.* **84** (1996) 1209

[12] J. Feldman, M. Salmhofer, and E. Trubowitz, *Comm. Pure Appl. Math.* **51** (1998) 1133, *Comm. Pure Appl. Math.* **52** (1999) 273, *Comm. Pure Appl. Math.* **53** (2000) 1350.

[13] J. Feldman, J. Magnen, V. Rivasseau, E. Trubowitz, *Helv.Phys.Acta* **65**(1992)679

[14] M. Salmhofer, *Rev. Math. Phys.* **10** (1998) 553

[15] J. Feldman, H. Knörrer, D. Lehmann, E. Trubowitz, in *Constructive Physics*, V. Rivasseau (ed.), Springer Lecture Notes in Physics, 1995

[16] N. Furukawa, T.M. Rice, M. Salmhofer, *Phys. Rev. Lett.* **81** (1998) 3195

[17] M. Disertori, V. Rivasseau, cond-mat/9907130

[18] D.C. Brydges, T. Kennedy, *J. Stat. Phys.* **48** (1987) 19

[19] M. Randeria et al., *Phys. Rev. Lett.* **74** (1995) 4951, J.C. Campuzano et al., *Phys. Rev.* **B52** (1995) 615