A rigorous study on the charged gap in strongly correlated electron clusters

2002 J. Phys. A: Math. Gen. 35 941

(http://iopscience.iop.org/0305-4470/35/4/306)

INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS A: MATHEMATICAL AND GENERAL

J. Phys. A: Math. Gen. **35** (2002) 941–950 PII: S0305-4470(02)24672-7

A rigorous study on the charged gap in strongly correlated electron clusters

Guang-Shan Tian and Ji-Gang Wang

Department of Physics, Peking University, Beijing 100871, People's Republic of China

Received 8 May 2001, in final form 5 December 2001
Published 18 January 2002
Online at stacks.iop.org/JPhysA/35/941

Abstract

In this paper, we study a revised definition $\tilde{\Delta}_c$ of the charged gap for the strongly correlated electron models on small clusters, proposed by Nishino, in a mathematically rigorous way. By applying a simplified version of Lieb's spin-reflection-positivity method, we show that this quantity is always positive for the half-filled Hubbard model, the periodic Anderson model and the Kondo lattice model. We also establish a model-dependent lower bound to the charged gap. Our results show explicitly the role played by electron repulsion in opening up a nonvanishing charged gap of a cluster.

PACS numbers: 71.10.Fd, 36.40.Cg, 71.27.+a

In the study of strongly correlated electron systems, the possible existence of charged gaps and spin gaps in their excitation spectrum attracts many physicists' interest. The quantum transport and the magnetic properties of these systems are mainly affected by these gaps. In particular, if a strongly correlated electron system has a nonzero charged gap at certain fillings, it becomes an insulator. For instance, in a seminal paper [1], Lieb and Wu solved exactly the one-dimensional Hubbard model and showed that, for any on-site Coulomb repulsion $U > 0$, the system has a nonvanishing charged gap at half-filling. Consequently, it is always a Mott insulator and the metal–insulator transition happens at $U_c = 0$ in this model. In higher dimensions, it is still debatable whether the half-filled Hubbard model has a metal–insulator transition at $U_c > 0$ [2, 3].

For some more complicated models, such as the periodic Anderson model and the Kondo lattice model, exact solutions are difficult to obtain. In these cases, physicists mainly rely on numerical calculations on small size samples to derive reliable results. For example, Tsunetsugu *et al* [4] studied the magnetic transition in a one-dimensional Kondo lattice by numerical diagonalization. They found that the spin gap is always nonzero when the system is half-filled. Therefore, the ground state of the Kondo chain is a spin liquid. After introduction of the density matrix renormalization group technique [5], Yu and White [6] could investigate the same model on a larger sample and confirmed the previous results. Moreover, these authors also observed that the charged gap of this model is larger than its spin excitation gap. A detailed

review on the recent progresses in study of the one-dimensional Kondo lattice chain can be found in [7].

In the above-mentioned works, the charged gap (or quasiparticle gap) of a strongly correlated electron model is defined by

$$\Delta_c \equiv E_0(N+1) + E_0(N-1) - 2E_0(N) \tag{1}$$

where $E_0(N)$ is the ground-state energy in the sector of N electrons (in particular, the subspace with $N = N_\Lambda$, the total number of lattice sites, is called the half-filled sector). However, we should notice that this definition has some shortcomings. First, it cannot be directly applied when an external magnetic field is turned on and a Zeeman interaction term is added to the Hamiltonian. Secondly, it is not suitable for numerical calculations on a small size sample with a dozen of lattice sites. In this case, the finite size effects caused by the discreteness of the single-particle levels makes Δ_c a strongly parity-dependent quantity. Namely, as N_Λ changes from an even integer to an odd integer, Δ_c fluctuates dramatically.

To overcome these problems, Nishino proposed an alternative definition of the charged gap [8]. Let $E_0(N_1, N_2)$ be the ground-state energy of the system in the subspace with N_1 up-spin and N_2 down-spin electrons. Nishino introduced

$$\tilde{\Delta}_c(n_\uparrow, n_\downarrow) \equiv E_0(n_\uparrow + 1, n_\downarrow + 1) + E_0(n_\uparrow, n_\downarrow) - E_0(n_\uparrow + 1, n_\downarrow) - E_0(n_\uparrow, n_\downarrow + 1) \tag{2}$$

for the charged gap of a strongly correlated model on a small cluster. In (2), $n_\uparrow$ and $n_\downarrow$ are subject to the constraint condition

$$n_\uparrow + n_\downarrow + 1 = N_\Lambda. \tag{3}$$

Otherwise, they may take on any admissible integer values.

By explicitly calculating $\tilde{\Delta}_c$ for an one-dimensional Hubbard cluster with three to ten sites, Nishino showed numerically that it is indeed much less parity dependent than Δ_c. He also pointed out that, for $n_\uparrow = n_\downarrow \sim N_\Lambda/2$ and $N_\Lambda \to \infty$, $\tilde{\Delta}_c$ tends to Δ_c. Furthermore, the effect of the Zeeman interaction on the charged gap can be easily taken into consideration by choosing a specific pair of integers $(n_\uparrow, n_\downarrow)$ with $2S_z = (n_\uparrow - n_\downarrow)$. Therefore, Nishino's definition for the charged gap is more practical for numerical calculations on small clusters.

Theoretically, a natural and interesting question that arises is whether, *for a strongly correlated electron cluster with repulsive interactions, $\tilde{\Delta}_c$ is always a non-negative quantity for any admissible pair of integers $(n_\uparrow, n_\downarrow)$, subject to condition (3).*

Recently, by applying a generalized version of Lieb's spin-reflection-positivity technique [9–11], we studied the excitation gaps in the Hubbard model, the periodic Anderson model, the Kondo lattice model and the double-exchange model [12–14]. We showed that, *in any dimensions*, the charged gaps (1) of these models at half-filling are larger than their spin excitation gaps. This conclusion confirms the previous results derived by numerical calculations [7].

In the present paper, we shall employ the same method, which will be further simplified in the following, to show that the charged gap $\tilde{\Delta}_c$ introduced by Nishino for a strongly correlated electron cluster is indeed a positive quantity. Our main results can be summarized in the following theorem.

Theorem. *For any pair of admissible integers $n_\uparrow$ and $n_\downarrow$ satisfying the constraint condition (3), the charged gap $\tilde{\Delta}_c(n_\uparrow, n_\downarrow)$ is a non-negative quantity, i.e.*

$$\tilde{\Delta}_c(n_\uparrow, n_\downarrow) \geqslant 0 \tag{4}$$

for the half-filled positive-U Hubbard model, the periodic Anderson model and the Kondo lattice model on a bipartite cluster. Moreover, for a specific model, this inequality can be

further strengthened. For example, for the half-filled positive-U Hubbard cluster, $\tilde{\Delta}_{\rm c}(n_\uparrow, n_\downarrow)$ *satisfies the following inequality*

$$\tilde{\Delta}_{\rm c}(n_\uparrow, n_\downarrow) \geqslant \frac{U}{N_\Lambda} \tag{5}$$

where U *is the Coulomb repulsion between electrons and* N_Λ *is the number of sites in the cluster.*

By this theorem, $\tilde{\Delta}_{\rm c}(n_\uparrow, n_\downarrow)$ is nonvanishing for any $U \neq 0$ and $n_\uparrow + n_\downarrow + 1 = N_\Lambda$. Therefore, it properly characterizes the charged gap caused by the Coulomb correlation between electrons in a small size cluster, even in an external magnetic field.

Before proceeding to the proof of this theorem, we would like to introduce some definitions and useful notation.

Take a finite d-dimensional cluster Λ with N_Λ sites. The Hamiltonian of the Hubbard model is of the following form:

$$H_{\rm H} = -t \sum_\sigma \sum_{\langle ij \rangle} \left(c^\dagger_{i\sigma} c_{j\sigma} + c^\dagger_{j\sigma} c_{i\sigma}\right) + U \sum_{i\in\Lambda} \left(n_{i\uparrow} - \tfrac{1}{2}\right)\left(n_{i\downarrow} - \tfrac{1}{2}\right) - \mu \hat{N}. \tag{6}$$

In (6), $c^\dagger_{i\sigma}$ ($c_{i\sigma}$) is the fermion creation (annihilation) operator which creates (annihilates) an *itinerant* electron of spin σ at site $\boldsymbol{i}$. $\langle \boldsymbol{ij} \rangle$ denotes a pair of cluster sites. The parameters $t > 0$ and $U > 0$ represent the kinetic energy and the on-site interaction between itinerant electrons, respectively. μ is the chemical potential. In terms of $H_{\rm H}$, Λ is bipartite if it can be divided into two separated sub-clusters A and B such that, electrons can only hop from a site in A (B) to a site in B (A). In the following, we shall only consider models on a bipartite cluster.

Similarly, the Hamiltonians of the Anderson model and the Kondo lattice model are defined by

$$\begin{aligned} H_{\rm A} = {} & -t \sum_\sigma \sum_{\langle ij \rangle} \left(c^\dagger_{i\sigma} c_{j\sigma} + c^\dagger_{j\sigma} c_{i\sigma}\right) + V \sum_\sigma \sum_{i\in\Lambda} \left(c^\dagger_{i\sigma} d_{i\sigma} + d^\dagger_{i\sigma} c_{i\sigma}\right) \\ & + U \sum_{i\in\Lambda} \left(n^d_{i\uparrow} - \tfrac{1}{2}\right)\left(n^d_{i\downarrow} - \tfrac{1}{2}\right) - \mu \hat{N} \end{aligned} \tag{7}$$

and

$$H_{\rm K} = -t \sum_\sigma \sum_{\langle ij \rangle} \left(c^\dagger_{i\sigma} c_{j\sigma} + c^\dagger_{j\sigma} c_{i\sigma}\right) + J \sum_{i\in\Lambda} \boldsymbol{\sigma}_i \cdot \boldsymbol{s}_i - \mu \hat{N}. \tag{8}$$

In (7), $c_{i\sigma}$ ($d_{i\sigma}$) represents the s(d)-orbital fermion operator at site $\boldsymbol{i}$. Similarly, in (8), $\boldsymbol{\sigma}_i$ and $\boldsymbol{s}_i$ denote the spin operators of itinerant electrons and localized electrons at site $\boldsymbol{i}$, respectively, and $J > 0$ is the antiferromagnetic super-exchange interaction between them. It is well known that these spin operators can be expressed by spin-$\frac{1}{2}$ fermion operators. For instance, we have

$$s_{ix} \equiv \tfrac{1}{2}\left(f^\dagger_{i\uparrow} f_{i\downarrow} + f^\dagger_{i\downarrow} f_{i\uparrow}\right) \qquad s_{iy} \equiv \frac{1}{2\mathrm{i}}\left(f^\dagger_{i\uparrow} f_{i\downarrow} - f^\dagger_{i\downarrow} f_{i\uparrow}\right) \qquad s_{iz} \equiv \tfrac{1}{2}\left(n^f_{i\uparrow} - n^f_{i\downarrow}\right) \tag{9}$$

where $f_{i\sigma}$ denotes the localized fermion operator at site $\boldsymbol{i}$. Since there is a localized spin at each site in the Kondo lattice model, these operators are subject to the following constraint condition

$$n^f_{\boldsymbol{i}} = f^\dagger_{i\uparrow} f_{i\uparrow} + f^\dagger_{i\downarrow} f_{i\downarrow} = 1. \tag{10}$$

For $H_{\rm A}$ and $H_{\rm K}$, the definition of a bipartite cluster is slightly complicated. A detailed discussion on this issue can be found in our previous paper [15].

These Hamiltonians enjoy some common symmetries, which are useful in our following analysis. First, these Hamiltonians commute with the total particle number operators $\hat{N}$.

Consequently, their Hilbert spaces can be divided into numerous subspaces $\{V(N)\}$. Each of these subspaces is characterized by an integer N, the total number of fermions in the system. In particular, the subspace $V(N = N_\Lambda)$ is called the half-filled subspace for the Hubbard Hamiltonian. However, for both the Anderson model and the Kondo lattice model, this subspace is $V(N = 2N_\Lambda)$.

Furthermore, it is easy to check that these Hamiltonians commute with the total spin operators $\hat{S}_+ = \hat{S}_x + \mathrm{i}\hat{S}_y$, $\hat{S}_- = \hat{S}_x - \mathrm{i}\hat{S}_y$ and $\hat{S}_z$. Therefore, S^2 and S_z are good quantum numbers of these systems.

Finally, it has been established that, when $\mu = 0$, *the global ground states* of these Hamiltonians on a bipartite cluster coincide with their ground states in the corresponding half-filled subspaces [9, 16]. Since we only consider the half-filled clusters in this paper, we shall set $\mu = 0$ in the following.

Now, we are ready to prove the theorem.

Proof of the theorem. In the following, we shall take the Hubbard model as a concrete example. The same technique can be easily applied to both the periodic Anderson model and the Kondo lattice model.

To begin with, let us consider the ground-state energy $E_0(n_\uparrow + 1, n_\downarrow + 1)$ of the Hubbard Hamiltonian with $n_\uparrow + n_\downarrow + 1 = N_\Lambda$. To apply the spin-reflection-positivity technique, we introduce a unitary partial particle–hole transformation $\hat{U}_{\mathrm{H}}$ [17–19][1], which is defined by

$$\hat{U}_{\mathrm{H}}^\dagger c_{i\uparrow}\hat{U}_{\mathrm{H}} = c_{i\uparrow} \qquad \hat{U}_{\mathrm{H}}^\dagger c_{i\downarrow}\hat{U}_{\mathrm{H}} = \epsilon(i)c_{i\downarrow}^\dagger \tag{11}$$

where $\epsilon(i) = 1$, for $i \in A$ and $\epsilon(i) = -1$, for $i \in B$. Under this transformation, the half-filled *positive-U* Hubbard Hamiltonian (with $\mu = 0$) is mapped into a *negative-U Hamiltonian of the same form*. In the meantime, the subspace $V(n_\uparrow + 1, n_\downarrow + 1)$ is mapped into $V(n_\uparrow + 1, N_\Lambda - n_\downarrow - 1) = V(n_\uparrow + 1, n_\uparrow)$. Consequently, we have the identity

$$E_0(n_\uparrow + 1, n_\downarrow + 1, U) = E_0(n_\uparrow + 1, n_\uparrow, -U) \tag{12}$$

where $E_0(n_\uparrow + 1, n_\uparrow, -U)$ denotes the ground-state energy of the negative-U Hubbard Hamiltonian in the subspace $V(n_\uparrow + 1, n_\uparrow)$.

Next, we construct the ground-state wavefunction $\Psi_0(n_\uparrow + 1, n_\uparrow, -U)$. Following [10], we define the quasi-fermion operators by

$$\hat{C}_{i\uparrow} \equiv c_{i\uparrow} \qquad \hat{C}_{i\downarrow} \equiv (-1)^{\hat{N}_\uparrow} c_{i\downarrow}. \tag{13}$$

In (13), $\hat{N}_\uparrow$ represents the total number operator of up-spin fermions in the system. It is easy to check that the conventional anticommutation relations

$$\{\hat{C}_{i\sigma}, \hat{C}_{j\sigma}\} = \{\hat{C}_{i\sigma}^\dagger, \hat{C}_{j\sigma}^\dagger\} = 0 \qquad \{\hat{C}_{i\sigma}^\dagger, \hat{C}_{j\sigma}\} = \delta_{ij} \tag{14}$$

still hold for operators with *the same spin indices*. However, operators $\{\hat{C}_{i\uparrow}\}$, now, *commute with* $\{\hat{C}_{i\downarrow}\}$. Consequently, the negative-U Hubbard Hamiltonian at half-filling (with $\mu = 0$) can be written as

$$H_{\mathrm{H}}(-U, \mu) = \hat{T}_\uparrow \otimes \hat{I}_\downarrow + \hat{I}_\uparrow \otimes \hat{T}_\downarrow - U \sum_{i\in\Lambda} \left(n_{i\uparrow} - \tfrac{1}{2}\right) \otimes \left(n_{i\downarrow} - \tfrac{1}{2}\right). \tag{15}$$

In (15), $\hat{T}_\sigma = -t \sum_{\langle ij\rangle} \left(\hat{C}_{i\sigma}^\dagger \hat{C}_{j\sigma} + \hat{C}_{j\sigma}^\dagger \hat{C}_{i\sigma}\right)$ is the hopping term of spin-σ fermions and $\hat{I}_\sigma$ represents the identity operator, acting in the Hilbert space of spin-σ fermions. Furthermore, $\Psi_0(n_\uparrow + 1, n_\uparrow, -U)$ can be written as a linear combination of $\{\psi_m^\uparrow \otimes \psi_n^\downarrow\}$, i.e.

$$\Psi_0(n_\uparrow + 1, n_\uparrow, -U) = \sum_{m,n} W_{mn}\psi_m^\uparrow \otimes \psi_n^\downarrow. \tag{16}$$

[1] For some applications of the partial particle–hole transformation in studying the spectral functions of the Hubbard model, see [19].

In (16), $\psi_{\rm k}^{\sigma}$ denotes a configuration of spin-σ fermions. More explicitly, it is defined by

$$\psi_{\rm k}^{\sigma} \equiv \hat{C}_{i_1\sigma}^{\dagger} \cdots \hat{C}_{i_L\sigma}^{\dagger} |0\rangle \tag{17}$$

where $(i_1, \ldots, i_L)$, with $L = n_{\uparrow} + 1$, for $\sigma = \uparrow$ and $L = n_{\uparrow}$, for $\sigma = \downarrow$, indicate the positions of fermions with spin σ in the lattice. $|0\rangle$ is the vacuum state. Apparently, the entire set $\{\psi_{\rm k}^{\sigma}\}$ gives a natural basis of $V_{\sigma}(L)$, the subspace of L quasi-fermions with spin σ.

By letting m be the row index and n be the column index, we can further write the coefficients $\{W_{mn}\}$ of $\Psi_0(n_{\uparrow} + 1, n_{\uparrow}, -U)$ into a matrix $\mathcal{W}$. But, in general, this matrix *is not a square matrix*. That is due to the fact that, for $\sigma = \uparrow$, $V_{\uparrow}(n_{\uparrow} + 1)$ has dimension $C_{N_\Lambda}^{n_{\uparrow}+1}$, while the dimension of $V_{\downarrow}(n_{\uparrow})$ is $C_{N_\Lambda}^{n_{\uparrow}}$. *Generally, these dimensions are not equal*. Mathematically, it is rather difficult to deal with a nonsquare matrix. In our previous papers [12, 13], we solved this problem by enlarging the Hilbert subspaces of the negative-U Hubbard Hamiltonian and then, constructing a new coefficient matrix $\widetilde{W}$, *which is square*. To $\widetilde{W}$, we were able to apply the standard polar decomposition theorem in matrix theory [20]. However, in this process, many unphysical states, *which are not eigenvectors of the particle number operators* $\hat{N}_{\uparrow}$ *and* $\hat{N}_{\downarrow}$, were created. To eliminate these states, we had to set their coefficients in the expansion of the wavefunction to be zero. That made our previous approach rather complicated.

In the following, we shall take a more straightforward approach, which is based on the following singular polar decomposition theorem for the nonsquare matrices in matrix theory.

Lemma (singular polar decomposition theorem). *Let A be an $m \times n$ matrix with $m \neq n$.*

(i) If $m < n$, then there exist an $m \times m$ unitary matrix U_1, an $m \times m$ diagonal semipositive definite matrix Λ_1 and an $m \times n$ matrix V_1 such that

$$A = U_1 \Lambda_1 V_1. \tag{18}$$

Moreover, the m rows of matrix V_1 are orthonormal vectors.

(ii) Similarly, if $m > n$, then there exist an $m \times n$ matrix V_2, an $n \times n$ diagonal semipositive definite matrix Λ_2 and an $n \times n$ unitary matrix U_2 such that

$$A = V_2 \Lambda_2 U_2 \tag{19}$$

with the n columns of V_2 being orthonormal.

The proof of this theorem can be found in a standard textbook on matrix theory [20]. For the reader's convenience, we shall give a short proof of it in the appendix of this paper.

For definiteness, let us assume that the coefficient matrix W has more rows than columns. In this case, the singular polar decomposition theorem tells us that there are three matrices U_W, V_W and Λ_W such that

$$W = V_W \Lambda_W U_W \tag{20}$$

where V_W is an $C_{N_\Lambda}^{n_{\uparrow}+1} \times C_{N_\Lambda}^{n_{\uparrow}}$ matrix with orthonormal columns and U_W is an $C_{N_\Lambda}^{n_{\uparrow}} \times C_{N_\Lambda}^{n_{\uparrow}}$ unitary matrix. Moreover, Λ_W is an $C_{N_\Lambda}^{n_{\uparrow}} \times C_{N_\Lambda}^{n_{\uparrow}}$ diagonal matrix with $\lambda_l \geqslant 0$. Consequently, the wavefunction $\Psi_0(n_{\uparrow} + 1, n_{\uparrow}, -U)$ can be rewritten as

$$\begin{aligned} \Psi_0(n_{\uparrow} + 1, n_{\uparrow}, -U) &= \sum_{m,n} W_{mn} \psi_m^{\uparrow} \otimes \psi_n^{\downarrow} = \sum_{m,n} (V_W \Lambda_W U_W)_{mn} \psi_m^{\uparrow} \otimes \psi_n^{\downarrow} \\ &= \sum_{l=1}^{D} \lambda_l \xi_l^{\uparrow} \otimes \phi_l^{\downarrow} \end{aligned} \tag{21}$$

with $D = C_{N_\Lambda}^{n_{\uparrow}}$. In (21), $\xi_l^{\uparrow}$ and $\phi_l^{\downarrow}$ are defined by

$$\xi_l^{\uparrow} = \sum_m (V_W)_{ml} \psi_m^{\uparrow} \qquad \phi_l^{\downarrow} = \sum_n (U_W)_{ln} \psi_n^{\downarrow}. \tag{22}$$

Since U_W is unitary and the columns of V_W are orthonormal, the new sets of vectors $\{\xi_l^{\uparrow}\}$ and $\{\phi_l^{\downarrow}\}$ are also orthonormal. More importantly, *these new vectors* $\{\xi_l^{\uparrow}\}$ *and* $\{\phi_l^{\downarrow}\}$ *are also the eigenvectors of the particle number operators* $\hat{N}_{\uparrow}$ *and* $\hat{N}_{\downarrow}$*, respectively*. Furthermore, because the ground-state wavefunction $\Psi_0(n_{\uparrow}+1, n_{\uparrow}, -U)$ is normalized, we have

$$\langle\Psi_0(n_{\uparrow}+1, n_{\uparrow}, -U)|\Psi_0(n_{\uparrow}+1, n_{\uparrow}, -U)\rangle = \mathrm{Tr}\, W^{\dagger}W = \sum_{l=1}^{D}\lambda_l^2 = 1. \tag{23}$$

With the simple form (21) of the wavefunction $\Psi_0(n_{\uparrow}+1, n_{\uparrow}, -U)$, we now calculate its energy $E_0(n_{\uparrow}+1, n_{\uparrow}, -U)$.

$$\begin{aligned}
&E_0(n_{\uparrow}+1, n_{\uparrow}, -U)\\
&\quad= \langle\Psi_0(n_{\uparrow}+1, n_{\uparrow}, -U)|H_{\mathrm{H}}(-U)|\Psi_0(n_{\uparrow}+1, n_{\uparrow}, -U)\rangle\\
&\quad= \sum_{l=1}^{D}\lambda_l^2\big[\langle\xi_l^{\uparrow}|\hat{T}_{\uparrow}|\xi_l^{\uparrow}\rangle + \langle\phi_l^{\downarrow}|\hat{T}_{\downarrow}|\phi_l^{\downarrow}\rangle\big]\\
&\qquad -U\sum_{i\in\Lambda}\Bigg(\sum_{l_1,l_2=1}^{D}\lambda_{l_1}\lambda_{l_2}\langle\xi_{l_2}^{\uparrow}|n_{i\uparrow}-\tfrac{1}{2}|\xi_{l_1}^{\uparrow}\rangle\langle\phi_{l_2}^{\downarrow}|n_{i\downarrow}-\tfrac{1}{2}|\phi_{l_1}^{\downarrow}\rangle\Bigg).
\end{aligned}\tag{24}$$

Now, without causing any confusion, we can drop the spin indices in (24). By applying inequality $|ab| \leqslant \frac{1}{2}(|a|^2+|b|^2)$ to each term in the last summation of (24), we obtain

$$\begin{aligned}
&E_0(n_{\uparrow}+1, n_{\uparrow}, -U)\\
&\quad\geqslant \tfrac{1}{2}\sum_{l=1}^{D}\lambda_l^2\big[\langle\xi_l|\hat{T}|\xi_l\rangle + \langle\xi_l|\hat{T}|\xi_l\rangle\big] + \tfrac{1}{2}\sum_{l=1}^{D}\lambda_l^2\big[\langle\phi_l|\hat{T}|\phi_l\rangle + \langle\phi_l|\hat{T}|\phi_l\rangle\big]\\
&\qquad -\frac{U}{2}\sum_{i\in\Lambda}\Bigg(\sum_{l_1,l_2=1}^{D}\lambda_{l_1}\lambda_{l_2}\langle\xi_{l_2}|n_i-\frac{1}{2}|\xi_{l_1}\rangle\overline{\langle\xi_{l_2}|n_i-\frac{1}{2}|\xi_{l_1}\rangle}\Bigg)\\
&\qquad -\frac{U}{2}\sum_{i\in\Lambda}\Bigg(\sum_{l_1,l_2=1}^{D}\lambda_{l_1}\lambda_{l_2}\langle\phi_{l_2}|n_i-\frac{1}{2}|\phi_{l_1}\rangle\overline{\langle\phi_{l_2}|n_i-\frac{1}{2}|\phi_{l_1}\rangle}\Bigg).
\end{aligned}\tag{25}$$

The right-hand side of (25) can be further simplified by introducing new wavefunctions

$$\Psi_1 = \sum_{l=1}^{D}\lambda_l\xi_l^{\uparrow}\otimes\bar{\xi}_l^{\downarrow} \qquad \Psi_2 = \sum_{l=1}^{D}\lambda_l\phi_l^{\uparrow}\otimes\bar{\phi}_l^{\downarrow} \tag{26}$$

where $\bar{\xi}_l^{\sigma}$ and $\bar{\phi}_l^{\sigma}$ are the complex conjugates of ξ_l^{σ} and ϕ_l^{σ}, respectively. In terms of Ψ_1 and Ψ_2, inequality (25) now reads

$$E_0(n_{\uparrow}+1, n_{\uparrow}, -U) \geqslant \tfrac{1}{2}\langle\Psi_1|H_{\mathrm{H}}(-U)|\Psi_1\rangle + \tfrac{1}{2}\langle\Psi_2|H_{\mathrm{H}}(-U)|\Psi_2\rangle. \tag{27}$$

Notice that, by their constructions, Ψ_1 and Ψ_2 are actually wavefunctions in the subspaces $V(n_{\uparrow}+1, n_{\uparrow}+1)$ and $V(n_{\uparrow}, n_{\uparrow})$, respectively. Therefore, by the variational principle, inequality (27) can be further reduced to

$$E_0(n_{\uparrow}+1, n_{\uparrow}, -U) \geqslant \tfrac{1}{2}E_0(n_{\uparrow}+1, n_{\uparrow}+1, -U) + \tfrac{1}{2}E_0(n_{\uparrow}, n_{\uparrow}, -U). \tag{28}$$

Under the inverse of the partial particle–hole transformation $\hat{U}_{\mathrm{H}}^{-1}$, (28) is transformed into the following equivalent inequality

$$E_0(n_{\uparrow}+1, n_{\downarrow}+1, U) \geqslant \tfrac{1}{2}E_0(n_{\uparrow}+1, n_{\downarrow}, U) + \tfrac{1}{2}E_0(n_{\uparrow}, n_{\downarrow}+1, U) \tag{29}$$

as we showed in (12). In inequality (29), $E_0(N_1, N_2, U)$ denotes the ground-state energy of the original *positive-U* Hubbard Hamiltonian in the subspace $V(N_1, N_2)$.

Similarly, by repeating the above steps, we can also show that

$$E_0(n_\uparrow, n_\downarrow, U) \geqslant \tfrac{1}{2}E_0(n_\uparrow, n_\downarrow + 1, U) + \tfrac{1}{2}E_0(n_\uparrow + 1, n_\downarrow, U) \tag{30}$$

holds true. Therefore, by summing up (29) and (30), we finally obtain

$$E_0(n_\uparrow + 1, n_\downarrow + 1, U) + E_0(n_\uparrow, n_\downarrow, U) \geqslant E_0(n_\uparrow + 1, n_\downarrow, U) + E_0(n_\uparrow, n_\downarrow + 1, U). \tag{31}$$

Namely, the charged gap $\tilde{\Delta}_c(n_\uparrow, n_\downarrow)$ defined in (2) is indeed a nonnegative quantity for any pair of admissible integers $n_\uparrow$ and $n_\downarrow$ subject to the constraint condition (3).

To establish the lower bound (5) to $\tilde{\Delta}_c(n_\uparrow, n_\downarrow)$, we now estimate the errors caused in deriving inequality (27). By replacing inequality $|ab| \leqslant |a|^2/2 + |b|^2/2$ with identity

$$a\bar{b} + \bar{a}b = |a|^2 + |b|^2 - |a - b|^2 \tag{32}$$

we can actually rewrite inequality (27) into an equation

$$\begin{aligned} E_0(n_\uparrow + 1, n_\uparrow, -U) &= \tfrac{1}{2}\langle\Psi_1|H_{\rm H}(-U)|\Psi_1\rangle + \tfrac{1}{2}\langle\Psi_2|H_{\rm H}(-U)|\Psi_2\rangle \\ &\quad + \frac{U}{2}\sum_{i\in\Lambda}\sum_{l_1,l_1}\lambda_{l_1}\lambda_{l_2}\left|\langle\xi_{l_2}|n_{i\uparrow} - \frac{1}{2}|\xi_{l_1}\rangle - \overline{\langle\phi_{l_2}|n_{i\downarrow} - \frac{1}{2}|\phi_{l_1}\rangle}\right|^2. \end{aligned} \tag{33}$$

Consequently, by the variational principle, we obtain an improved inequality

$$\begin{aligned} &E_0(n_\uparrow + 1, n_\uparrow, -U) - \tfrac{1}{2}E_0(n_\uparrow + 1, n_\uparrow + 1, -U) - \tfrac{1}{2}E_0(n_\uparrow, n_\uparrow, -U) \\ &\quad \geqslant \frac{U}{2}\sum_{i\in\Lambda}\sum_{l_1,l_1}\lambda_{l_1}\lambda_{l_2}\left|\langle\xi_{l_2}|n_i - \frac{1}{2}|\xi_{l_1}\rangle - \overline{\langle\phi_{l_2}|n_i - \frac{1}{2}|\phi_{l_1}\rangle}\right|^2 \\ &\quad \geqslant \frac{U}{2}\sum_{i\in\Lambda}\sum_{l}\lambda_l^2\left(\langle\xi_l|n_i - \frac{1}{2}|\xi_l\rangle - \langle\phi_l|n_i - \frac{1}{2}|\phi_l\rangle\right)^2. \end{aligned} \tag{34}$$

The right-hand side of (34) can be further simplified by applying the Cauchy–Schwartz inequality $|\sum a_n b_n|^2 \leqslant \sum |a_n|^2 \sum |b_n|^2$. We have

$$\begin{aligned} &\frac{U}{2}\sum_{i\in\Lambda}\sum_{l}\lambda_l^2\left(\langle\xi_l|n_i - \frac{1}{2}|\xi_l\rangle - \langle\phi_l|n_i - \frac{1}{2}|\phi_l\rangle\right)^2 \\ &\quad \geqslant \frac{U}{2}\left[\sum_{i\in\Lambda}\sum_{l}\lambda_l^2\right]^{-1}\left[\sum_{i\in\Lambda}\sum_{l}\lambda_l^2(\langle\xi_l|n_i - \tfrac{1}{2}|\xi_l\rangle - \langle\phi_l|n_i - \tfrac{1}{2}|\phi_l\rangle)\right]^2 \\ &\quad = \frac{U}{2N_\Lambda}\langle\Psi_0(n_\uparrow + 1, n_\uparrow, -U)|(\hat{N}_\uparrow - \hat{N}_\downarrow)|\Psi_0(n_\uparrow + 1, n_\downarrow, -U)\rangle^2 \\ &\quad = \frac{U}{2N_\Lambda}. \end{aligned} \tag{35}$$

As a result, inequality (29) is strengthened and can be rewritten as

$$\Delta_1 \equiv E_0(n_\uparrow + 1, n_\downarrow + 1, U) - \frac{1}{2}E_0(n_\uparrow + 1, n_\downarrow, U) - \frac{1}{2}E_0(n_\uparrow, n_\downarrow + 1, U) \geqslant \frac{U}{2N_\Lambda}. \tag{36}$$

Similarly, the same lower bound can also be established to the energy difference

$$\Delta_2 \equiv E_0(n_\uparrow, n_\downarrow, U) - \frac{1}{2}E_0(n_\uparrow + 1, n_\downarrow, U) - \frac{1}{2}E_0(n_\uparrow, n_\downarrow + 1, U) \geqslant \frac{U}{2N_\Lambda}. \tag{37}$$

Therefore, the charged gap $\tilde{\Delta}_c = \Delta_1 + \Delta_2$ is bounded below by U/N_Λ.

Our theorem is proven. □

Some remarks are in order.

Remark 1. In the above, we take the Hubbard Hamiltonian as a concrete example. In fact, the proof can be easily extended to the periodic Anderson model, another strongly correlated electron model without localized spins. However, for the Kondo lattice model, some technical problems may arise, due to the constraint condition (10) on the localized spins.

As shown above, in the first step of proof, we need to introduce a proper partial particle–hole transformation U_{K} and map the Kondo lattice into a unitarily equivalent Hamiltonian $\tilde{H}_{\mathrm{K}}$ with negative coupling constants [16]. Under the same transformation, the constraint condition (10) now reads

$$\hat{f}_{i\uparrow}^{\dagger}\hat{f}_{i\uparrow} = \hat{f}_{i\downarrow}^{\dagger}\hat{f}_{i\downarrow}. \tag{38}$$

It requires that, at each lattice site, there must be the same number of up-spin and down-spin fermions. It is this constraint condition which makes the construction of the ground-state wavefunction of $\tilde{H}_{\mathrm{K}}$ rather difficult. To deal with this problem, one needs to decompose the subspace $V(n_{\uparrow}+1, n_{\uparrow}, -J)$ of the transformed Hamiltonian $\tilde{H}_{\mathrm{K}}$ into numerous smaller subspaces and pick up the ones which are subject to condition (38). A detailed discussion on these subtle points can be found in our previous papers [12–14].

Remark 2. In the second part of our proof, we established a lower bound to $\tilde{\Delta}_{\mathrm{c}}(n_{\uparrow}, n_{\downarrow})$. However, this bound tends to zero as $N_{\Lambda} \to \infty$. Therefore, it cannot tell us whether a nonvanishing charged gap sustains in the thermodynamic limit. However, (33), the main equation in establishing the lower bound to $\tilde{\Delta}_{\mathrm{c}}$ gives us a strong hint that this question may be addressed by deriving more detailed information on the ground-state wavefunction $\widetilde{\Psi}_0(n_{\uparrow}+1, n_{\uparrow}, -U)$ of the *negative-U* Hubbard Hamiltonian, by numerical calculations. A definite conclusion could also be reached by applying more sophisticated methods than just applying the Cauchy–Schwartz inequality. Further investigations in this direction will be carried out in future.

In summary, in this paper, we investigate a revised definition $\tilde{\Delta}_{\mathrm{c}}$ of the charged gap for the strongly correlated electron models on small clusters, proposed by Nishino [8]. By applying a simplified version of Lieb's spin-reflection-positivity method, we prove that this quantity is always positive for the half-filled Hubbard model, the periodic Anderson model and the Kondo lattice model. We also establish a model-dependent lower bound to the charged gap. Our results show explicitly the role played by the electron repulsions in opening up a nonvanishing charged gap in a small cluster.

Acknowledgments

The authors would like to thank Dr T Nishino for sending us [8]. They would also like to thank Dr T Xiang for useful discussions. This work was partially supported by the Chinese National Science Foundation under grant no 19874004.

Appendix

In this appendix, for the reader's convenience, we shall give a brief proof of the singular polar decomposition theorem, which we applied to prove the main theorem of this paper. One can find a more detailed proof of this theorem in [20].

Proof of the singular polar decomposition theorem. First, let us assume that $m < n$. In this case, we consider matrix product $AA^{\dagger}$. It is an $m \times m$ semipositive definite matrix. Therefore, it has m orthonormal eigenvectors $(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m)$, which satisfy equations

$$AA^{\dagger}\boldsymbol{x}_i = \lambda_i^2 \boldsymbol{x}_i,\ 1 \leqslant i \leqslant m. \tag{39}$$

Re-organizing $\{\lambda_i\}$ in decreasing order

$$\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_{\mathrm{K}} > \lambda_{k+1} = \cdots = \lambda_m = 0 \tag{40}$$

we define a diagonal semipositive definite matrix Λ_1 and an $m \times m$ unitary matrix U_1 by

$$\Lambda_1 = \begin{pmatrix} \lambda_1 \cdots 0 \\ \cdots\cdots \\ 0 \cdots \lambda_m \end{pmatrix} \tag{41}$$

and

$$U_1 = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m) \tag{42}$$

where $\boldsymbol{x}_i$ represents the ith column of matrix U_1.

Next, we construct matrix V_1. The first k rows of V_1 are given by

$$V_1^i = \frac{1}{\lambda_i} \left(A^\dagger \boldsymbol{x}_i\right)^\dagger. \tag{43}$$

Since $\lambda_i \neq 0$ for $1 \leqslant i \leqslant k$, these rows are well defined. To define the other $m - k$ rows of V_1, we notice that *the first k rows defined in (43) are orthonormal to each other*. Actually, we have

$$\begin{aligned} \overline{\langle V_1^i | V_1^j \rangle} &= \frac{1}{\lambda_i \lambda_j} \langle A^\dagger \boldsymbol{x}_i | A^\dagger \boldsymbol{x}_j \rangle \\ &= \frac{1}{\lambda_i \lambda_j} \boldsymbol{x}_i^\dagger A A^\dagger \boldsymbol{x}_j = \frac{1}{\lambda_i \lambda_j} \lambda_j^2 \boldsymbol{x}_i^\dagger \boldsymbol{x}_j = \delta_{ij}. \end{aligned} \tag{44}$$

In the last step of the above derivation, we used the definition of the vectors $\{\boldsymbol{x}_i\}$. On the other hand, since each row V_1^i is an n-dimensional vector, one can find other $m - k$ orthonormal vectors $\boldsymbol{z}_1^\dagger, \boldsymbol{z}_2^\dagger, \ldots, \boldsymbol{z}_{m-k}^\dagger$, which are orthogonal to each V_1^i with $1 \leqslant i \leqslant k$. We let them be the remaining $m - k$ rows of V_1. Consequently, matrix V_1 has m orthonormal rows.

Finally, we need to show that

$$U_1^\dagger A = \Lambda_1 V_1 \tag{45}$$

holds for the above defined matrices. Obviously, by their definitions, the first k rows of $U_1^\dagger A$ and $\Lambda_1 V_1$ are correspondingly identical. Consequently, we need only to consider the remaining $m - k$ rows of both $U_1^\dagger A$ and $\Lambda_1 V_1$. For $\Lambda_1 V_1$, these rows are zero vectors since $\lambda_{k+1} = \cdots = \lambda_m = 0$. We now show that the corresponding rows in $U_1^\dagger A$ are also zero vectors.

Let us take one row $\boldsymbol{x}_l^\dagger A$ of $U_1^\dagger A$ with $k + 1 \leqslant l \leqslant m$ and calculate its norm.

$$\begin{aligned} \langle \boldsymbol{x}_l^\dagger A | \boldsymbol{x}_l^\dagger A \rangle &= \sum_{\alpha=1}^{n} \overline{\left(\sum_{\beta=1}^{m} \overline{(\boldsymbol{x}_l)_\beta} A_{\beta\alpha} \right)} \left(\sum_{\gamma=1}^{m} \overline{(\boldsymbol{x}_l)_\gamma} A_{\gamma\alpha} \right) \\ &= \sum_{\alpha=1}^{n} \left[\sum_{\beta=1}^{m} \sum_{\gamma=1}^{m} \overline{A_{\beta\alpha}} (\boldsymbol{x}_l)_\beta A_{\gamma\alpha} \overline{(\boldsymbol{x}_l)_\gamma} \right] \\ &= \sum_{\alpha=1}^{n} \left[\sum_{\beta=1}^{m} \sum_{\gamma=1}^{m} \overline{(\boldsymbol{x}_l)_\gamma} A_{\gamma\alpha} A_{\alpha\beta}^\dagger (\boldsymbol{x}_l)_\beta \right] \\ &= \boldsymbol{x}_l^\dagger A A^\dagger \boldsymbol{x}_l = \lambda_l^2 \boldsymbol{x}_l^\dagger \boldsymbol{x}_l = 0. \end{aligned} \tag{46}$$

Therefore, $\boldsymbol{x}_l^\dagger A = 0$ and (45) is an identity.

Similarly, when $m > n$, one can prove that $A = V_2 \Lambda_2 U_2$ holds by considering the matrix product $A^\dagger A$ instead of $A A^\dagger$. □

References

[1] Lieb E H and Wu F Y 1968 *Phys. Rev. Lett.* **20** 1445
[2] Schrieffer J R, Wen X G and Zhang S C 1988 *Phys. Rev. Lett.* **60** 944
[3] Georges A, Kotliar G, Krauth W and Rozenberg M J 1966 *Rev. Mod. Phys.* **68** 13
[4] Tsunetsugu H, Hatsugai Y, Ueda K and Sigrist M 1992 *Phys. Rev.* B **46** 3175
[5] White S R 1993 *Phys. Rev. Lett.* **69** 2863
White S R 1993 *Phys. Rev.* B **48** 10 345
[6] Yu C C and White S R 1993 *Phys. Rev. Lett.* **71** 3866
[7] Tsunetsugu H, Sigrist M and Ueda K 1997 *Rev. Mod. Phys.* **69** 809 and references therein
[8] Nishino T 1993 *Physica* B **186–8** 885
[9] Lieb E H 1989 *Phys. Rev. Lett.* **62** 1201
[10] Lieb E H and Nachtergaele B 1995 *Phys. Rev.* B **51** 4777
[11] Tian G S 1997 *J. Phys. A: Math. Gen.* **30** 5329
[12] Tian G S 1998 *Phys. Rev.* B **58** 7612
[13] Tian G S and Tang L H 1999 *Phys. Rev.* B **60** 11 336
[14] Wang J G and Tian G S 2000 *Commun. Theor. Phys.* **34** 21
[15] Tian G S 1994 *Phys. Rev.* B **50** 6246
[16] Yanagisawa T and Shimoi Y 1995 *Phys. Rev. Lett.* **74** 4939
Tsunetsugu H 1997 *Phys. Rev.* B **55** 3042
[17] Shiba H 1972 *Prog. Theor. Phys.* **48** 2171
[18] Yang C N and Zhang S C 1990 *Mod. Phys. Lett.* B **4** 759
[19] Tian G S 1997 *Phys. Lett.* A **228** 383
[20] Horn R A and Johnson C R 1985 *Matrix Analysis* (Cambridge: Cambridge University Press)