

## DENSITY MATRIX MINIMIZATION WITH $\ell_1$ REGULARIZATION\*

RONGJIE LAI<sup>†</sup>, JIANFENG LU<sup>‡</sup>, AND STANLEY OSHER<sup>§</sup>

**Abstract.** We propose a convex variational principle to find sparse representation of low-lying eigenspace of symmetric matrices. In the context of electronic structure calculation, this corresponds to a sparse density matrix minimization algorithm with  $\ell_1$  regularization. The minimization problem can be efficiently solved by a split Bregman iteration type algorithm. We further prove that from any initial condition, the algorithm converges to a minimizer of the variational principle.

**Key words.** Density matrix,  $\ell_1$  regularization, eigenspace.

**AMS subject classifications.** 65K10, 65F30.

### 1. Introduction

The low-lying eigenspace of operators has many important applications, including those in quantum chemistry, numerical PDEs, and statistics. Given a  $n \times n$  symmetric matrix  $H$ , and denote its eigenvectors as  $\{\Phi_i\}, i = 1, \dots, n$ . The low-lying eigenspace is given by the span of the first  $N$  (usually  $N \ll n$ ) eigenvectors.

In many scenario, the real interest is the subspace itself, but not a particular set of basis functions. In particular, we are interested in a sparse representation of the eigenspace. The eigenvectors form a natural basis set, but oftentimes they are not sparse or localized (consider for example the eigenfunctions of the free Laplacian operator  $-\Delta$  on a periodic box). This suggests asking for an alternative sparse representation of the eigenspace.

In quantum chemistry, the low-lying eigenspace for a Hamiltonian operator corresponds to the physically occupied space of electrons. In this context, a localized class of basis functions of the low-lying eigenspaces is called Wannier functions [28, 14]. These functions provide transparent interpretation and understandings of covalent bonds, polarizations, *etc.* of the electronic structure. These localized representations are also the starting point and the essence for many efficient algorithms for electronic structure calculations (see e.g. the review article [10]).

**1.1. Our contribution.** In this work, we propose a convex minimization principle for finding a sparse representation of the low-lying eigenspace.

$$\begin{aligned} \min_{P \in \mathbb{R}^{n \times n}} \quad & \operatorname{tr}(HP) + \frac{1}{\mu} \|P\|_1 \\ \text{s.t.} \quad & P = P^T, \operatorname{tr} P = N, 0 \preceq P \preceq I, \end{aligned} \tag{1.1}$$

where  $\|\cdot\|_1$  is the entry-wise  $\ell_1$  matrix norm,  $A \preceq B$  denotes that  $B - A$  is a positive semi-definite matrix, and  $\mu$  is a penalty parameter for entry-wise sparsity. Here  $H$  is an

---

\*Received: March 7, 2014; accepted (in revised form): October 28, 2014. Communicated by Shi Jin. The research of J.L. was supported in part by the Alfred P. Sloan Foundation and the National Science Foundation under award DMS-1312659. The research of S.O. was supported by the Office of Naval Research Grant N00014-11-1-719. We thank the anonymous referees for helpful comments and suggestions.

<sup>†</sup>Department of Mathematics, Rensselaer Polytechnic Institute, 110 8th St, Troy, NY 12180, U.S.A. (lair@rpi.edu).

<sup>‡</sup>Departments of Mathematics, Physics, and Chemistry, Duke University, Durham, NC 27708, U.S.A. (jianfeng@math.duke.edu).

<sup>§</sup>Department of Mathematics and Institute for Pure and Applied Mathematics, University of California, Los Angeles, Los Angeles, CA 90095, U.S.A. (sjo@math.ucla.edu).

$n \times n$  symmetric matrix, which is the (discrete) Hamiltonian in the electronic structure context. The variational principle gives  $P$  as a sparse representation of the projection operator onto the low-lying eigenspace.

The key observation here is to use the matrix  $P$  instead of the wave functions  $\Psi$ . This leads to a convex variational principle. Physically, this corresponds to looking for a sparse representation of the density matrix. We also noted that in cases where we expect degeneracy or near-degeneracy of eigenvalues of the matrix  $H$ , the formulation in terms of the density matrix  $P$  is more natural, as it allows fractional occupation of states. This is a further advantage besides the convexity.

Moreover, we design an efficient minimization algorithm based on split Bregman iteration to solve the above variational problem. Starting from any initial condition, the algorithm always converges to a minimizer. In addition, the new variational problem (1.1) enables a promising way of designing efficient algorithms for finding a sparse density matrix, which would lead to new linear scaling algorithms in the field of electronic structure computation. This will be investigated in our future work.

**1.2. Previous works.** There is an enormous amount literature on numerical algorithms for Wannier functions and more generally sparse representation of low-lying eigenspace. The influential work [20] proposed a minimization strategy within the occupied space to find spatially localized Wannier functions (coined as “maximally localized Wannier functions”).

In [5], the second author with his collaborators developed a localized subspace iteration (LSI) algorithm to find Wannier functions. The idea behind the LSI algorithm is to combine the localization step with the subspace iteration method as an iterative algorithm to find Wannier functions of an operator. The method has been applied to electronic structure calculation in [9]. As [9] shows, due to the truncation step involved, the LSI algorithm does not in general guarantee convergence.

As a more recent work in [25],  $L_1$  regularization is proposed to be used in the variational formulation of the Schrödinger equation of quantum mechanics for creating compressed modes, a set of spatially localized functions  $\{\psi_i\}_{i=1}^N$  in  $\mathbb{R}^d$  with compact support.

$$E = \min_{\Psi_N} \sum_{j=1}^N \left( \frac{1}{\mu} |\psi_j|_1 + \langle \psi_j, \hat{H} \psi_j \rangle \right) \quad \text{s.t.} \quad \langle \psi_j, \psi_k \rangle = \delta_{jk}, \quad (1.2)$$

where  $\hat{H} = -\frac{1}{2}\Delta + V(x)$  is the Hamilton operator corresponding to potential  $V(x)$ , and the  $L_1$  norm is defined as  $|\psi_j|_1 = \int |\psi_j| dx$ . This  $L_1$  regularized variational approach describes a general formalism for obtaining localized (in fact, compactly supported) solutions to a class of mathematical physics PDEs, which can be recast as variational optimization problems. Although an efficient algorithm based on a method of splitting orthogonality constraints (SOC) [17] is designed to solve the above non-convex problem, it is still a big challenge to theoretically analyze the convergence of the proposed the algorithm.

The key idea in the proposed convex formulation (1.1) of the variational principle is the use of the density matrix  $P$ . The density matrix is widely used in electronic structure calculations, for example the density matrix minimization algorithm [18]. In this type of algorithm, sparsity of density matrix is specified explicitly by restricting the matrix to be a banded matrix. The resulting minimization problem is then non-convex and found to suffer from many local minimizers. Other electronic structure algorithms that

use density matrix include density matrix purification [21], Fermi operator expansion algorithm [1], just to name a few.

From a mathematical point of view, the use of density matrix can be viewed as similar to the idea of lifting, which has been recently used in recovery problems [2]. While a nuclear norm is used in the PhaseLift method [2] to enhance sparsity in terms of matrix rank; we will use an entry-wise  $\ell_1$  norm to favor sparsity in matrix entries.

The rest of the paper is organized as follows. We formulate and explain the convex variational principle for finding localized representations of the low-lying eigenspace in Section 2. An efficient algorithm is proposed in Section 3 to solve the variational principle, with numerical examples presented in Section 4. The convergence proof of the algorithm is given in Section 5.

## 2. Formulation

Let us denote by  $H$  a symmetric matrix<sup>1</sup> coming from, for example, the discretization of an effective Hamiltonian operator in electronic structure theory. We are interested in a sparse representation of the eigenspace corresponding to its low-lying eigenvalues. In physical applications, this corresponds to the occupied space of a Hamiltonian; in data analysis, this corresponds to the principal components (for which we take the negative of the matrix so that the largest eigenvalue becomes the smallest). We are mainly interested in physics application here, and henceforth, we will mainly interpret the formulation and algorithms from a physical view point.

The Wannier functions, originally defined for periodic Schrödinger operators, are spatially localized basis functions of the occupied space. In [25], it was proposed to find the spatially localized functions by minimizing the variational problem

$$\min_{\Psi \in \mathbb{R}^{n \times N}, \Psi^T \Psi = I} \text{tr}(\Psi^T H \Psi) + \frac{1}{\mu} \|\Psi\|_1 \quad (2.1)$$

where  $\|\Psi\|_1$  denotes the entry-wise  $\ell_1$  norm of  $\Psi$ . Here  $N$  is the number of Wannier functions and  $n$  is the number of spatial degree of freedom (e.g. number of spatial grid points or basis functions).

The idea of the above minimization can be easily understood by looking at each term in the energy functional. The  $\text{tr}(\Psi^T H \Psi)$  is the sum of the Ritz value in the space spanned by the columns of  $\Psi$ . Hence, without the  $\ell_1$  penalty term, the minimization

$$\min_{\Psi \in \mathbb{R}^{n \times N}, \Psi^T \Psi = I} \text{tr}(\Psi^T H \Psi) \quad (2.2)$$

gives the eigenspace corresponds to the first  $N$  eigenvalues (here and below, we assume the non-degeneracy that the  $N$ -th and  $(N+1)$ -th eigenvalues of  $H$  are different). While the  $\ell_1$  penalty prefers  $\Psi$  to be a set of sparse vectors. The competition of the two terms gives a sparse representation of a subspace that is close to the eigenspace.

Due to the orthonormality constraint  $\Psi^T \Psi = I$ , the minimization problem (2.1) is not convex, which may result in troubles in finding the minimizer of the above minimization problem and also makes the proof of convergence difficult.

Here we take an alternative viewpoint, which gives a convex optimization problem. The key idea is instead of  $\Psi$ , we consider  $P = \Psi \Psi^T \in \mathbb{R}^{n \times n}$ . Since the columns of  $\Psi$  form an orthonormal set of vectors,  $P$  is the projection operator onto the space spanned by  $\Psi$ . In physical terms, if  $\Psi$  are the eigenfunctions of  $H$ ,  $P$  is then the density matrix which corresponds to the Hamiltonian operator. For insulating systems,

<sup>1</sup>With obvious changes, our results generalize to the Hermitian case.

it is known that the off-diagonal terms in the density matrix decay exponentially fast [3, 4, 6, 7, 13, 15, 22, 23, 26].

We propose to look for a sparse approximation of the exact density matrix by solving the minimization problem proposed in (1.1). The variational problem (1.1) is a convex relaxation of the non-convex variational problem

$$\begin{aligned} \min_{P \in \mathbb{R}^{n \times n}} \quad & \text{tr}(HP) + \frac{1}{\mu} \|P\|_1 \\ \text{s.t.} \quad & P = P^T, \text{tr} P = N, P = P^2, \end{aligned} \quad (2.3)$$

where the constraint  $0 \preceq P \preceq I$  is replaced by the idempotency constraint of  $P$ :  $P = P^2$ . The variational principle (2.3) can be understood as a reformulation of (2.1) using the density matrix as variable. The idempotency condition  $P = P^2$  is indeed the analog of the orthogonality constraint  $\Psi^T \Psi = I$ . Note that  $0 \preceq P \preceq I$  requires that the eigenvalues of  $P$  (the occupation number in physical terms) are between 0 and 1, while  $P = P^2$  requires the eigenvalues are either 0 or 1. Hence, the set

$$\mathcal{C} = \{P : P = P^T, \text{tr} P = N, 0 \preceq P \preceq I\} \quad (2.4)$$

is the convex hull of the set

$$\mathcal{D} = \{P : P = P^T, \text{tr} P = N, P = P^2\}. \quad (2.5)$$

Therefore (1.1) is indeed a convex relaxation of (2.3).

Without the  $\ell_1$  regularization, the variational problems (1.1) and (2.3) become

$$\begin{aligned} \min_{P \in \mathbb{R}^{n \times n}} \quad & \text{tr}(HP) \\ \text{s.t.} \quad & P = P^T, \text{tr} P = N, 0 \preceq P \preceq I, \end{aligned} \quad (2.6)$$

and

$$\begin{aligned} \min_{P \in \mathbb{R}^{n \times n}} \quad & \text{tr}(HP) \\ \text{s.t.} \quad & P = P^T, \text{tr} P = N, P = P^2. \end{aligned} \quad (2.7)$$

These two minimizations actually lead to the same result in the non-degenerate case.

**PROPOSITION 2.1.** *Let  $H$  be a symmetric  $n \times n$  matrix. Assume that the  $N$ -th and  $(N+1)$ -th eigenvalues of  $H$  are distinct, the minimizers of (2.6) and (2.7) are the same.*

This is perhaps a folklore result in linear algebra, nevertheless we include the short proof here for completeness.

*Proof.* It is clear that the unique minimizer of (2.7) is given by the projection matrix on the first  $N$  eigenvectors of  $H$ , given by

$$P_N = \sum_{i=1}^N v_i v_i^T$$

where  $v_i$  for  $i = 1, \dots, n$  are the eigenvectors of  $H$ , ordered according to their associated eigenvalues. Let us prove that (2.6) is minimized by the same solution.

Assume  $P$  is a minimizer of (2.6), we calculate

$$\text{tr}(HP) = \sum_{i=1}^n v_i^T H P v_i = \sum_{i=1}^n \lambda_i v_i^T P v_i = \sum_{i=1}^n \lambda_i \theta_i(P), \tag{2.8}$$

where  $\theta_i(P) = v_i^T P v_i$ . On the other hand, we have

$$\text{tr}(P) = \sum_{i=1}^n v_i^T P v_i = \sum_{i=1}^n \theta_i(P) = N,$$

and  $0 \leq \theta_i(P) \leq 1$  since  $0 \preceq P \preceq I$ . Therefore, if we view (2.8) as a variational problem with respect to  $\{\theta_i\}$ , it is clear that the unique minimum is achieved when

$$\theta_i(P) = \begin{cases} 1, & i \leq N; \\ 0, & \text{otherwise.} \end{cases}$$

We conclude the proof by noticing that the above holds if and only if  $P = P_N$ . □

This result states that we can convexify the set of admissible matrices. We remark that, somewhat surprisingly, this result also holds for the Hartree–Fock theory [19] which can be vaguely understood as a nonlinear eigenvalue problem. However the resulting variational problem is still non-convex for the Hartree–Fock theory.

Proposition 2.1 implies that the variational principle (1.1) can be understood as an  $\ell_1$  regularized version of the variational problem (2.7). The equivalence no longer holds for (1.1) and (2.3) with the  $\ell_1$  regularization. The advantage of (1.1) over (2.3) is that the former is a convex problem while the latter is not.

Coming back to the properties of the variational problem (1.1). We note that while the objective function of (1.1) is convex, it is not strictly convex as the  $\ell_1$  norm is not strictly convex and the trace term is linear. Therefore, in general, the minimizer of (1.1) is not unique.

EXAMPLE 2.1. *Let  $\mu \in \mathbb{R}_+$ ,  $N = 1$  and*

$$H = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \tag{2.9}$$

*The non-uniqueness comes from the degeneracy of the Hamiltonian eigenvalues. Any diagonal matrix  $P$  with trace 1 and non-negative diagonal entries is a minimizer.*

EXAMPLE 2.2. *Let  $\mu = 1$ ,  $N = 1$  and*

$$H = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 2 \\ 0 & 2 & 2 \end{pmatrix} \tag{2.10}$$

*The non-uniqueness comes from the competition between the trace term and the  $\ell_1$  regularization. The eigenvalues of  $H$  are 0, 1, and 4. A straightforward calculation shows that*

$$P_0 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1/2 & -1/2 \\ 0 & -1/2 & 1/2 \end{pmatrix} \tag{2.11}$$

which corresponds to the eigenvector  $(0, \sqrt{2}/2, -\sqrt{2}/2)^T$  associated with eigenvalue 0 and

$$P_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{2.12}$$

which corresponds to the eigenvector  $(1, 0, 0)^T$  associated with eigenvalue 1 are both minimizers of the objective function  $\text{tr}(HP) + \|P\|_1$ . Actually, due to convexity, any convex combination of  $P_0$  and  $P_1$  is a minimizer too.

It is an open problem under what assumptions that the uniqueness is guaranteed.

### 3. Algorithm

To solve the proposed minimization problem (1.1), we design an algorithm based on split Bregman iteration [12], which comes from the ideas of variables splitting and Bregman iteration [24]. Bregman iteration has attained intensive attention due to its efficiency in many  $\ell_1$  related constrained optimization problems [31, 30]. With the help of auxiliary variables, split Bregman iteration iteratively approaches the original optimization problem by computation of several easy-to-solve subproblems. This algorithm popularizes the idea of using operator/variable splitting to solve optimization problems arising from information science. The equivalence of the split Bregman iteration to the alternating direction method of multipliers (ADMM), Douglas–Rachford splitting and augmented Lagrangian method can be found in [8, 27, 29].

By introducing auxiliary variables  $Q = P$  and  $R = P$ , the optimization problem (1.1) is equivalent to

$$\begin{aligned} \min_{P, Q, R \in \mathbb{R}^{n \times n}} \quad & \frac{1}{\mu} \|Q\|_1 + \text{tr}(HP) \\ \text{s.t.} \quad & Q = P, R = P, \text{tr} P = N, R = R^T, 0 \preceq R \preceq I, \end{aligned} \tag{3.1}$$

which can be iteratively solved by:

$$(P^k, Q^k, R^k) = \arg \min_{P, Q, R \in \mathbb{R}^{n \times n}} \frac{1}{\mu} \|Q\|_1 + \text{tr}(HP) + \frac{\lambda}{2} \|P - Q + b\|_F^2 + \frac{r}{2} \|P - R + d\|_F^2 \tag{3.2}$$

$$\text{s.t.} \quad \text{tr} P = N, R = R^T, 0 \preceq R \preceq I,$$

$$b^k = b^{k-1} + P^k - Q^k \tag{3.3}$$

$$d^k = d^{k-1} + P^k - R^k \tag{3.4}$$

where variables  $b, d$  are essentially Lagrangian multipliers and parameters  $r, \lambda$  control the penalty terms. Solving  $P^k, Q^k, R^k$  in (3.3) alternatively, we have the following algorithm.

**Algorithm 3.1.** Initialize  $Q^0 = R^0 = P^0 \in \mathcal{C}, b^0 = d^0 = 0$

**while** “not converge” **do**

1.  $P^k = \arg \min_{P \in \mathbb{R}^{n \times n}} \operatorname{tr}(HP) + \frac{\lambda}{2} \|P - Q^{k-1} + b^{k-1}\|_F^2 + \frac{r}{2} \|P - R^{k-1} + d^{k-1}\|_F^2, \text{ s.t. } \operatorname{tr} P = N.$
2.  $Q^k = \arg \min_{Q \in \mathbb{R}^{n \times n}} \frac{1}{\mu} \|Q\|_1 + \frac{\lambda}{2} \|P^k - Q + b^{k-1}\|_F^2.$
3.  $R^k = \arg \min_{R \in \mathbb{R}^{n \times n}} \frac{r}{2} \|P^k - R + d^{k-1}\|_F^2, \text{ s.t. } R = R^T, 0 \preceq R \preceq I.$
4.  $b^k = b^{k-1} + P^k - Q^k.$
5.  $d^k = d^{k-1} + P^k - R^k.$

Note that the minimization problems in the steps of Algorithm 3.1 can be solved explicitly, as follows:

$$P^k = B^k - \frac{\operatorname{tr}(B^k) - N}{n}, \tag{3.5}$$

$$\text{where } B^k = \frac{\lambda}{\lambda+r} (Q^{k-1} - b^{k-1}) + \frac{r}{\lambda+r} (R^{k-1} - d^{k-1}) - \frac{1}{\lambda+r} H$$

$$Q^k = \operatorname{Shrink} \left( P^k + b^{k-1}, \frac{1}{\lambda\mu} \right) = \operatorname{sign}(P^k + b^{k-1}) \max \left\{ |P^k + b^{k-1}| - \frac{1}{\lambda\mu}, 0 \right\} \tag{3.6}$$

$$R^k = V \min\{\max\{D, 0\}, 1\} V^T, \text{ where } [V, D] = \operatorname{eig}(P^k + d^{k-1}). \tag{3.7}$$

Starting from any initial guess, the following theorem guarantees that the algorithm converges to one of the minimizers of the variational problem (3.1).

**THEOREM 3.2.** *The sequence  $\{(P^k, Q^k, R^k)\}_k$  generated by Algorithm 3.1 from any starting point converges to a minimum of the variational problem (3.1).*

We will prove a slightly more general version of the above (Theorem 5.4). The idea of the proof follows from the standard framework of analyzing split Bregman iteration, *i.e.* alternating direction method of multipliers (ADMM), see for example [11]. The standard proof needs to be generalized to cover the current case of non-uniqueness due to non-strictly convexity of the functionals. We defer the proof to Section 5.

Note that the Algorithm 3.1 has a cubic scaling computational cost with respect to the dimension of the matrix  $H$ , since the step (3) involves a singular value decomposition for eigenvalue thresholding. As the final minimizer of the problem is expected to be localized, it is possible to design a linear scaling algorithm by exploiting the sparsity of the matrices, along the line of density matrix minimization methods for electronic structure calculations (see [18]). This is our ongoing research.

#### 4. Numerical results

In this section, numerical experiments are presented to demonstrate the proposed model (1.1) for density matrix computation using Algorithm 3.1. We illustrate our numerical results in three representative cases, free electron model, Hamiltonian with energy band gap and a non-uniqueness example of the proposed optimization problem. All numerical experiments are implemented by MATLAB in a PC with a 16G RAM and a 2.7 GHz CPU.

**4.1. 1D Laplacian.** In the first example, we consider the proposed model for the free electron case, in other words, we consider the potential free Schrödinger operator  $-1/2\Delta$  defined on 1D domain  $\Omega = [0, 100]$  with periodic boundary condition.

This model approximates the behavior of valence electrons in a metallic solid with weak atomic pseudopotentials. In this case, the matrix  $H$  is a central difference discretization of  $-1/2\Delta$  on  $[0, 100]$  with equally spaced 256 points, and we take  $N = 10$ . Figure 4.1(a) illustrates the true density matrix  $\sum_{i=1}^{10} |\phi_i\rangle\langle\phi_i|$  obtained by the first 10 eigenfunctions of  $H$ . As the free Laplacian does not have a spectral gap, the density matrix decays slowly in the off-diagonal direction. Figure 4.1(b) and (c) plot the density matrices obtained from the proposed model with parameter  $\mu = 10$  and 100. Note that they are much localized than the original density matrix. As  $\mu$  gets larger, the variational problem imposes a smaller penalty on the sparsity, and hence the solution for  $\mu = 100$  has a wider spread than that for  $\mu = 10$ .

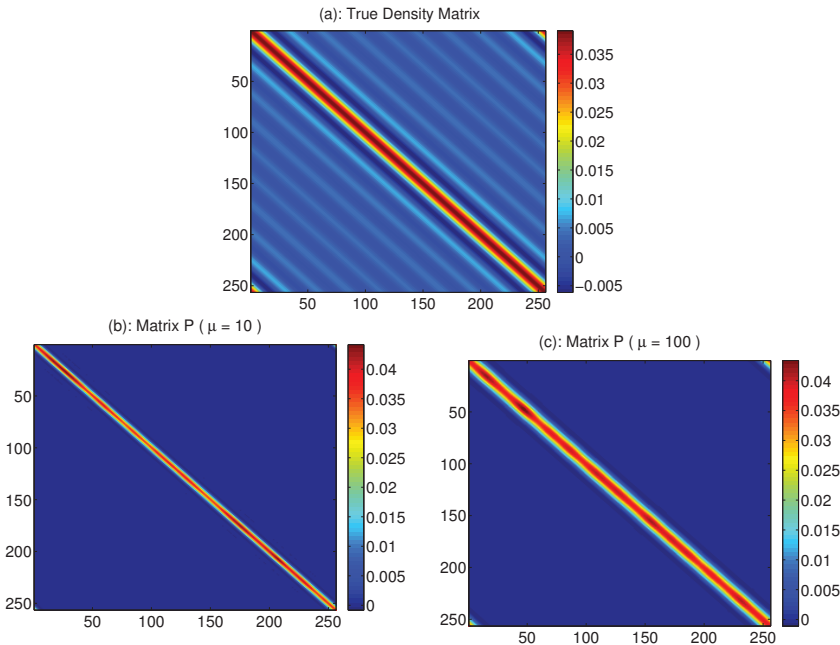


FIG. 4.1. (a) The true density matrix obtained by the first 10 eigenfunctions of  $H$ . (b), (c) solutions of the density matrices with  $\mu = 10, 100$  respectively.

After we obtain the sparse representation of the density matrix  $P$ , we can find localized Wannier functions as its action on the delta functions, as plotted in Figure 4.2 upper and lower pictures for  $\mu = 10$  and 100 respectively.

To indicate the approximation behavior of the proposed model, we consider the energy function approximation of  $\frac{1}{\mu}\|P\|_1 + \text{tr}(HP)$  to  $\sum_{i=1}^{10} \langle\phi_i|H|\phi_i\rangle$  with different values of  $\mu$ . In addition, we define  $\sum_{i=1}^{10} \|\phi_i - P\phi_i\|^2$  as a measurement for the space approximation of the density matrix  $P$  to the lower eigenspace  $\text{Span}\{\phi_i\}_{i=1}^{10}$ . Figure 4.3 reports the energy approximation and the space approximation with different values of  $\mu$ . Both numerical results suggest that the proposed model will converge to the energy states of the Schrödinger operator. We also remark that even though the exact density matrix is not sparse, a sparse approximation gives fairly good results in terms of energy and space approximations.



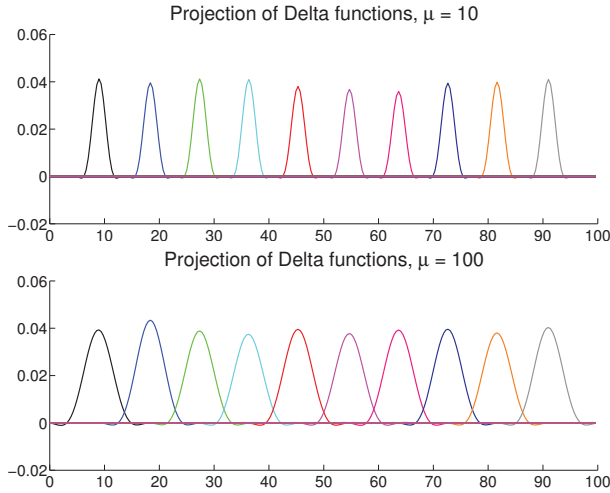


FIG. 4.2. Projection of Delta function  $\delta(x - x_i)$  using density matrices with  $\mu = 10$  (upper) and  $\mu = 100$  (lower) respectively.

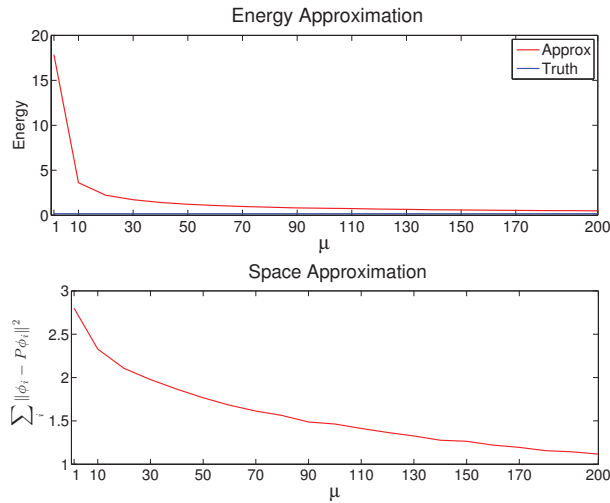


FIG. 4.3. Upper: Energy approximation as a function of  $\mu$ . Lower: Space approximation as a function of  $\mu$ .

**4.2. 1D Hamiltonian operator with a band gap.** We now consider a modified Kronig–Penney (KP) model [16] for a one-dimensional insulator. The original KP model describes the states of independent electrons in a one-dimensional crystal, where the potential function  $V(x)$  consists of a periodic array of rectangular potential wells. We replace the rectangular wells with inverted Gaussians so that the potential is given by

$$V(x) = -V_0 \sum_{j=1}^{N_{\text{at}}} \exp \left[ -\frac{(x - x_j)^2}{\delta^2} \right],$$

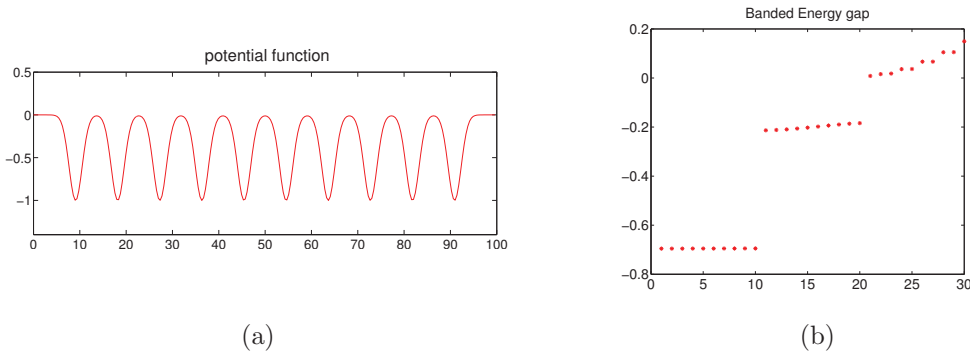


FIG. 4.4. (a) The potential function in the modified Kronig–Penney model. (b) The spectrum of the (discretized) Hamiltonian operator.

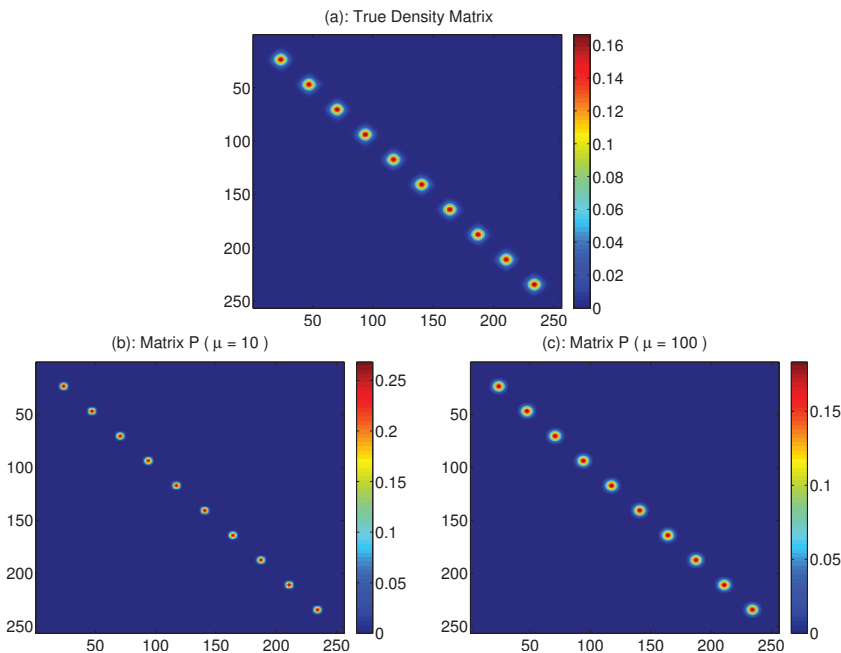


FIG. 4.5. (a) The true density matrix obtained by the first 10 eigenfunctions of  $H$ . (b), (c) solutions of the density matrices with  $\mu = 10, 100$  respectively.

where  $N_{\text{at}}$  gives the number of potential wells. In our numerical experiments, we choose  $N_{\text{at}} = 10$  and  $x_j = 100j/11$  for  $j = 1, \dots, N_{\text{at}}$ , and the domain is  $[0, 100]$  with periodic boundary condition. The potential is plotted in Figure 4.4(a). For this given potential, the Hamiltonian operator  $H = -\frac{1}{2}\Delta + V(x)$  exhibits two low-energy bands separated by finite gaps from the rest of the eigenvalue spectrum (See Figure 4.4(b)). Here a centered difference is used to discretize the Hamiltonian operator.

We consider three choices of  $N$  for this model:  $N = 10$ ,  $N = 15$  and  $N = 20$ . They correspond to three interesting physical situations of the model, as explained below.

For  $N = 10$ , the first band of the Hamiltonian is occupied, and hence the system has a spectral gap between the occupied and unoccupied states. As a result, the associated density matrix is exponentially localized, as shown in Figure 4.5(a). The resulting

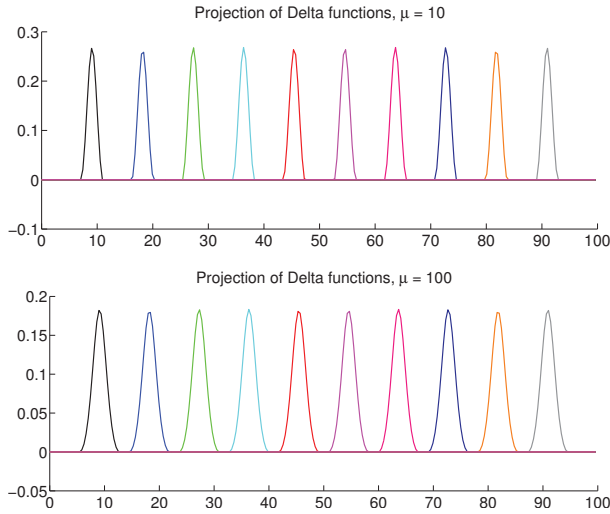


FIG. 4.6. Projection of Delta function  $\delta(x - x_i)$  using density matrices with  $\mu = 10$  (upper) and  $\mu = 100$  (lower) respectively.

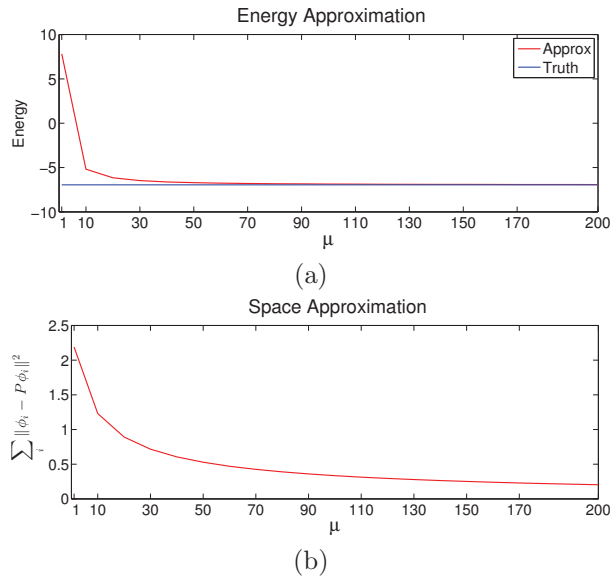


FIG. 4.7. (a) Energy approximation as a function of  $\mu$ . (b) Space approximation as a function of  $\mu$ .

sparse representation from the convex optimization is shown in Figure 4.5(b) and (c) for  $\mu = 10$  and 100 respectively. We see that the sparse representation agrees well with the exact density matrix, as the latter is very localized. The Wannier functions obtained by projection of delta functions are shown in Figure 4.6. As the system is an insulator, we see that the localized representation converges quickly to the exact answer when  $\mu$  increases. This is further confirmed in Figure 4.7 where the energy corresponding to the approximated density matrix and space approximation measurement  $\sum_{i=1}^{10} \|\phi_i - P\phi_i\|^2$  are plotted as functions of  $\mu$ .

Next we consider the case  $N = 15$ . The first band of 10 eigenstates of  $H$  is occupied

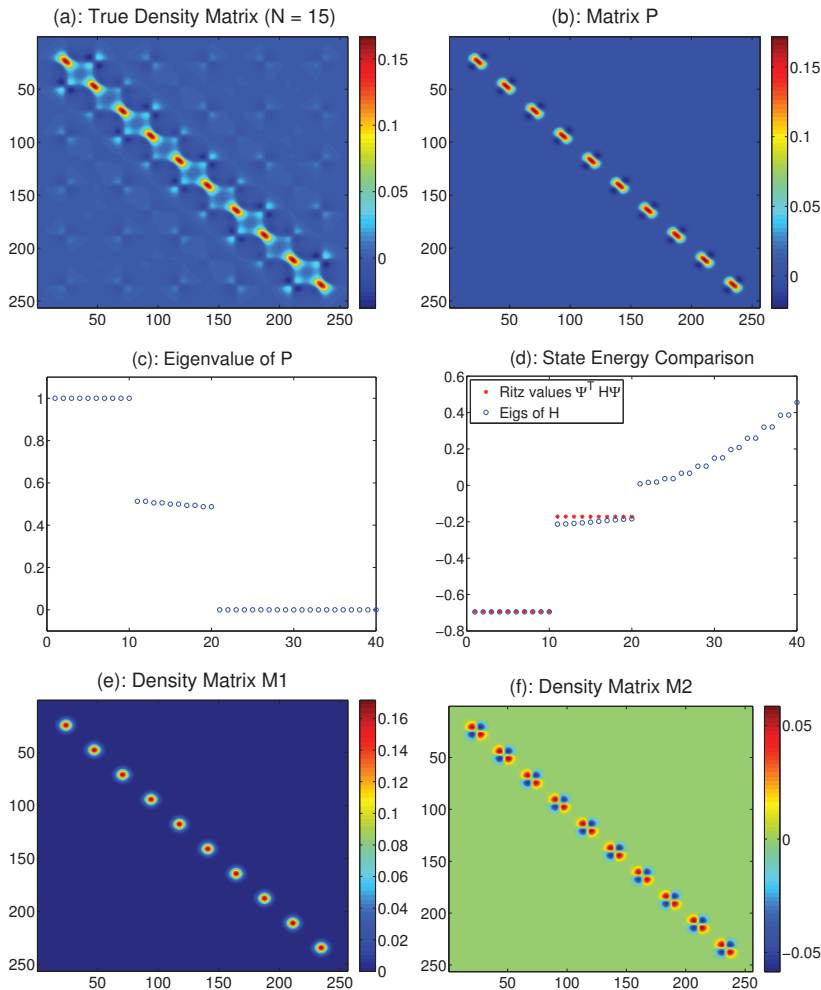


FIG. 4.8. (a) The true density matrix corresponds to the first 15 eigenfunctions of  $H$ . (b) The sparse representation  $P$  of the density matrix for  $\mu = 100$ . (c) The occupation number (eigenvalues) of  $P$ . (d) The first 15 eigenvalues of  $PH$  compared with the eigenvalues of  $H$ . (e) The filtered density matrix  $M_1$  corresponds to the first 10 eigenstates of  $P$ . (f) The filtered density matrix  $M_2$  corresponds to the next 10 eigenstates of  $P$ .

and the second band of  $H$  is “half-filled”. That is we have only 5 electrons occupying the 10 eigenstates of comparable eigenvalue of  $H$ . Hence, the system does not have a gap, which is indicated by the slow decay of the density matrix shown in Figure 4.8(a). Nevertheless, the algorithm with  $\mu = 100$  gives a sparse representation of the density matrix, which captures the feature of the density matrix near the diagonal, as shown in Figure 4.8(b). To understand better the resulting sparse representation, we diagonalize the matrix  $P$ :

$$P = \sum_i f_i \varphi_i \varphi_i^T.$$

The eigenvalues  $f_i$ , known as the occupation number in the physics literature, are sorted in the decreasing order. The first 40 occupation numbers are shown in Figure 4.8(c). We

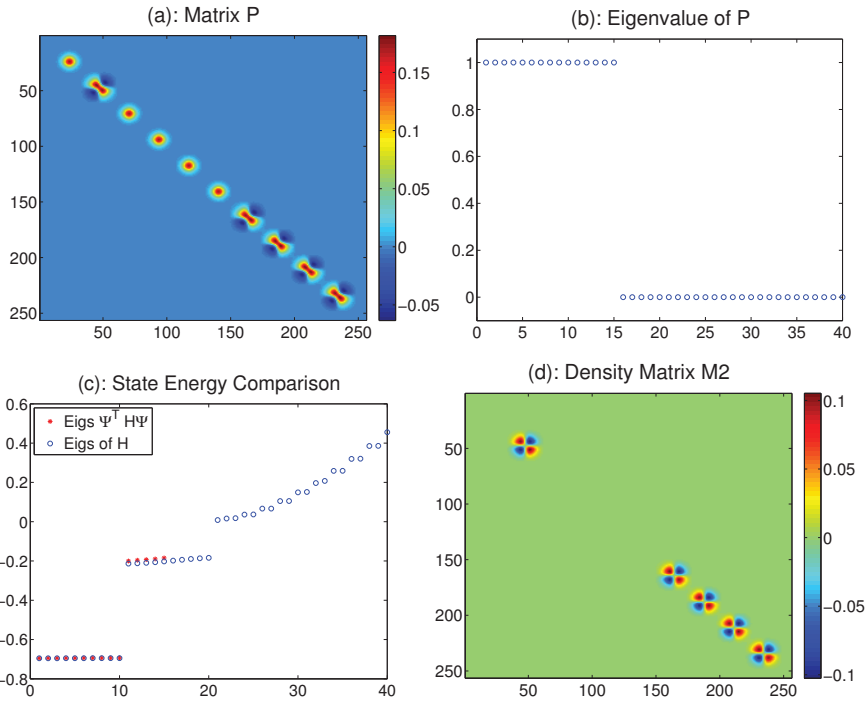


FIG. 4.9. Results obtained by the first 15 Compressed modes  $\Psi = \{\psi_i\}_{i=1}^{15}$  for  $\mu = 100$ . (a) The density representation  $P$  given by  $P = \Psi^T \Psi$ . (b) The occupation number (eigenvalues) of  $P$ . (c) The first 15 eigenvalues of  $\Psi^T H \Psi$  compared with the eigenvalues of  $H$ . (d) The filtered density matrix  $M_2$  corresponds to the 5 states in the second band.

have  $\sum_i f_i = \text{tr } P = 15$ , and we see that  $\{f_i\}$  exhibits two groups. The first 10 occupation numbers are equal to 1, corresponding to the fact that the lowest 10 eigenstates of the Hamiltonian operator is occupied. Indeed, if we compare the eigenvalues of the operator  $PH$  with the eigenvalues of  $H$ , as in Figure 4.8(d), we see that the first 10 low-lying states are well represented in  $P$ . This is further confirmed by the filtered density matrix  $M_1$  given by the first 10 eigenstates of  $P$  as

$$M_1 = \sum_{i=1}^{10} f_i \varphi_i \varphi_i^T,$$

plotted in Figure 4.8(e). It is clear that it is very close to the exact density matrix corresponding to the first 10 eigenfunctions of  $H$ , as plotted in Figure 4.5(a). The next group of occupation numbers in Figure 4.8(c) gets value close to 0.5. This indicates that those states are “half-occupied”, matches very well with the physical intuition. This is also confirmed by the state energy shown in Figure 4.8(d). Note that due to the fact these states are half filled, the perturbation in the eigenvalues by the localization is much stronger. The corresponding filtered density matrix

$$M_2 = \sum_{i=11}^{20} f_i \varphi_i \varphi_i^T,$$

is shown in Figure 4.8(f).

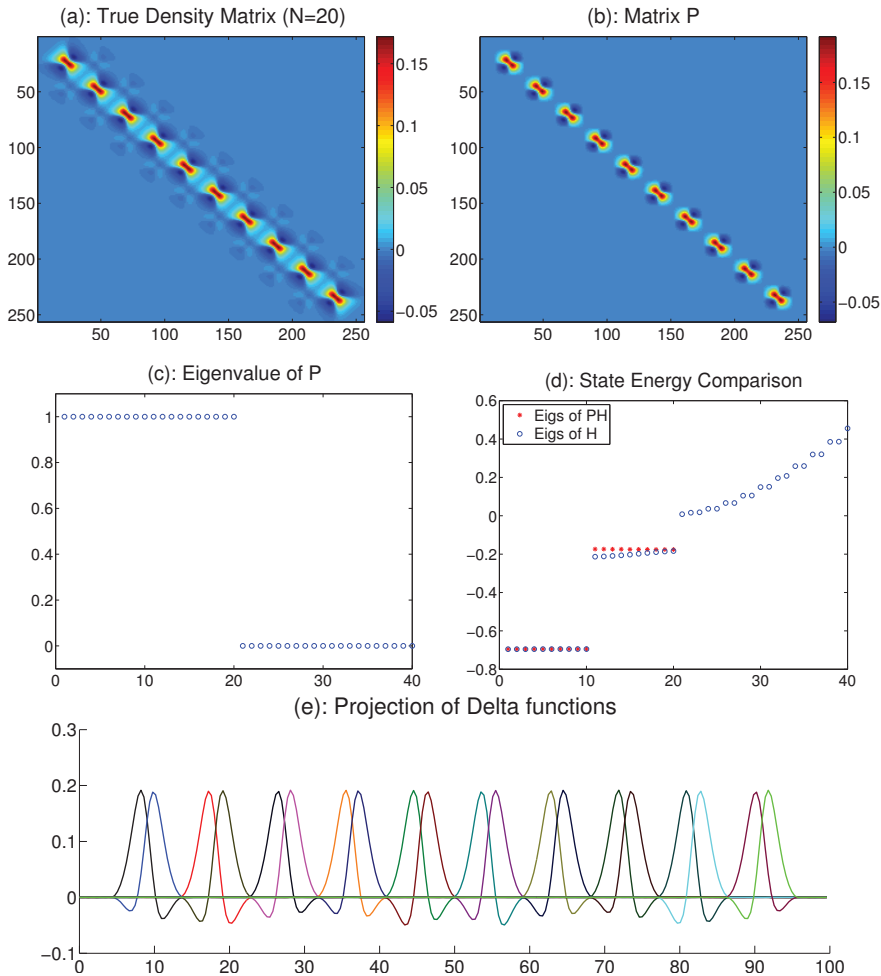


FIG. 4.10. (a) The true density matrix corresponds to the first 20 eigenfunctions of  $H$ . (b) The sparse representation  $P$  of the density matrix for  $\mu = 100$ . (c) The occupation number (eigenvalues) of  $P$ . (d) The first 20 eigenvalues of  $PH$  compared with the eigenvalues of  $H$ . (e) Projection of Delta function  $\delta(x - x_i)$ .

For this example, we compare with the results obtained using the variational principle (2.1) as in [25] shown in Figure 4.9. As the variational principle (2.1) is formulated with orbital functions  $\Psi$ , it does not allow fractional occupations, in contrast with the one in terms of the density matrix. Hence, the occupation number is either 1 or 0, which is equivalent to the idempotency condition, as shown in Figure 4.9(b). As a result, even though the states in the second band have very similar energy, the resulting  $\Psi$  are forced to choose five states over the ten, as can be seen from the Ritz value plotted in Figure 4.9(c). The solution is quite degenerate in this case. Physically, what happens is that the five electrons choose 5 wells out of the ten to sit in (on top of the state corresponding to the first band already in the well), as shown from the corresponding density matrix in Figure 4.9(a), or more clearly by the filtered density matrix in Figure 4.9(d) for the five higher energy states.

Finally, the  $N = 20$  case corresponds to the physical situation that the first two bands are all occupied. Note that as the band gap between the second band from

the rest of the spectrum is smaller than the gap between the first two bands; the density matrix, while still exponentially localized, has a slower off diagonal decay rate. The exact density matrix corresponds to the first 20 eigenfunctions of  $H$  is shown in Figure 4.10(a), and the localized representation with  $\mu=100$  is given in Figure 4.10(b). The occupation number plotted in Figure 4.10(c) indicates that the first 20 states are fully occupied, while the rest of the states are empty. This is further confirmed by comparison of the eigenvalues given by  $HP$  and  $H$ , shown in Figure 4.10(d). In this case, we see that physically, each well contains two states. Hence, if we look at the electron density, which is diagonal of the density matrix, we see a double peak in each well. Using the projection of delta functions, we see that the sparse representation of the density matrix  $P$  automatically locate the two localized orbitals centered at the two peaks, as shown in Figure 4.10(e).

**4.3. An example of non-unique minimizers.** Let us revisit the Example 2 in Section 2 for which the minimizers to the variational problem is non-unique. Theorem 3.2 guarantees that the algorithm will converge to some minimizer starting from any initial condition.

It is easy to check that in this case

$$P^* = Q^* = R^* = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad b^* = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & -1 & 1 \end{pmatrix}, \quad d^* = \begin{pmatrix} 0 & 0 & 0 \\ 0 & -1 & -1 \\ 0 & -1 & -1 \end{pmatrix} \quad (4.1)$$

is a fixed point of the algorithm. In Figure 4.11, we plot the sequence  $\left\{ \lambda \|b^k - b^*\|^2 + r \|d^k - d^*\|^2 + \lambda \|Q^k - Q^*\|^2 + r \|R^k - R^*\|^2 \right\}_k$  for a randomly chosen initial data. We see that the distance does not converge to 0 as the algorithm converges to another minimizer of the variational problem. Nonetheless, as will be shown in the proof of Theorem 3.2 in Section 5, the sequence is monotonically non-increasing.

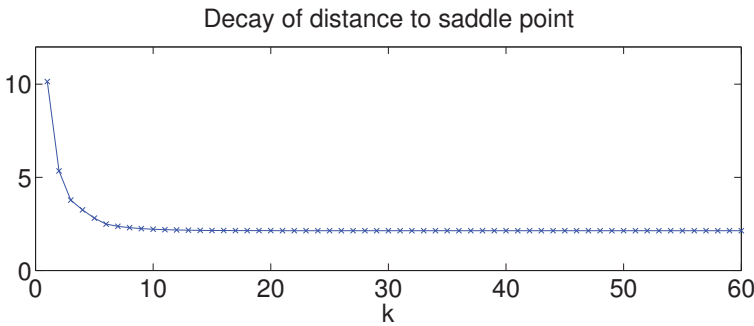


FIG. 4.11.  $\lambda \|b^k - b^*\|^2 + r \|d^k - d^*\|^2 + \lambda \|Q^k - Q^*\|^2 + r \|R^k - R^*\|^2$  as a function of  $k$  for Algorithm 3.1.

**5. Convergence of Algorithm 3.1**

For ease of notation, we will prove the convergence of the algorithm for the following slightly generalized variational problem:

$$\begin{aligned} & \min_{P,Q,R} f(P) + g(Q) + h(R) \\ & \text{s.t. } P = Q, \quad \text{and } P = R \end{aligned} \quad (5.1)$$

where  $f, g,$  and  $h$  are proper convex functionals, but not necessarily strictly convex. In particular, we will get (3.1) if we set

$$\begin{aligned}
 f(P) &= \begin{cases} \text{tr}(HP), & \text{if } \text{tr } P = N; \\ +\infty, & \text{otherwise,} \end{cases} \\
 g(Q) &= \|Q\|_1/\mu \\
 h(R) &= \begin{cases} 0, & \text{if } R = R^T, \text{ and } 0 \preceq R \preceq I; \\ +\infty, & \text{otherwise.} \end{cases}
 \end{aligned}$$

The corresponding algorithm for (5.1) is given by

**Algorithm 5.1.** *Initialize*  $P^0 = Q^0 = R^0, b^0 = d^0 = 0$   
**while** “not converge” **do**

- 1.  $P^k = \arg \min_P f(P) + \frac{\lambda}{2} \|P - Q^{k-1} + b^{k-1}\|^2 + \frac{r}{2} \|P - R^{k-1} + d^{k-1}\|^2,$
- 2.  $Q^k = \arg \min_Q g(Q) + \frac{\lambda}{2} \|P^k - Q + b^{k-1}\|^2.$
- 3.  $R^k = \arg \min_R h(R) + \frac{r}{2} \|P^k - R + d^{k-1}\|^2.$
- 4.  $b^k = b^{k-1} + P^k - Q^k.$
- 5.  $d^k = d^{k-1} + P^k - R^k.$

REMARK 5.1. Observe that the steps (2) and (3) above are independent and can be combined as a single minimization of the pair  $(Q^k, R^k)$ . Hence, even though the algorithm appears to be a “two level splitting”, viewing in terms of  $P$  and  $(Q, R)$ , it can be understood as a standard ADMM algorithm. Our analysis of the algorithm follows from the standard framework in [11], slightly generalized to deal with the possible non-uniqueness of the minimizer. We include the proof here to make the paper self-contained.

We define an augmented Lagrangian

$$\begin{aligned}
 \mathcal{L}(P, Q, R; b, d) &= f(P) + g(Q) + h(R) + \frac{\lambda}{2} \|P - Q\|^2 + \lambda \langle P - Q, b \rangle \\
 &\quad + \frac{r}{2} \|P - R\|^2 + r \langle P - R, d \rangle. \tag{5.2}
 \end{aligned}$$

DEFINITION 5.2. We call  $(P^*, Q^*, R^*; b^*, d^*)$  a saddle point of the Lagrangian (5.2), if

$$\mathcal{L}(P^*, Q^*, R^*; b, d) \leq \mathcal{L}(P^*, Q^*, R^*; b^*, d^*) \leq \mathcal{L}(P, Q, R; b^*, d^*) \tag{5.3}$$

for any  $(P, Q, R; b, d) \in \mathbb{R}^{n \times n} \times \mathbb{R}^{n \times n} \times \mathbb{R}^{n \times n} \times \mathbb{R}^{n \times n} \times \mathbb{R}^{n \times n}.$

LEMMA 5.3.  $(P^*, Q^*, R^*)$  is a solution of the optimization problem (5.1) if and only if there exist  $b^*, d^* \in \mathbb{R}^{n \times n}$  such that  $(P^*, Q^*, R^*; b^*, d^*)$  is a saddle point satisfying (5.3)

*Proof.* Given a saddle point  $(P^*, Q^*, R^*; b^*, d^*)$  satisfying (5.3), it is clear that the first inequality in (5.3) implies  $P^* = Q^* = R^*$ . Substitute  $P = Q = R$  in the second inequality of (5.3), we can immediately have  $(P^*, Q^*, R^*)$  is a minimizer (5.1).



On the other hand, suppose  $(P^*, Q^*, R^*)$  is a solution of (5.1). The first inequality in (5.3) holds since  $P^* = Q^* = R^*$ . Moreover, there exist  $b^*$  and  $d^*$  such that

$$-\lambda b^* - rd^* \in \partial f(P^*), \quad \lambda b^* \in \partial g(Q^*), \quad rd^* \in \partial h(R^*)$$

which suggests, for any  $P, Q, R \in \mathbb{R}^{n \times n}$

$$\begin{aligned} f(P) &\leq f(P^*) + \lambda \langle b^*, P - P^* \rangle + r \langle d^*, P - P^* \rangle \\ g(Q) &\leq g(Q^*) - \lambda \langle b^*, Q - Q^* \rangle \\ h(R) &\leq h(R^*) - r \langle d^*, R - R^* \rangle \end{aligned}$$

The summation of the above three inequalities yield the second inequality in (5.3).  $\square$

**THEOREM 5.4.** *The sequence  $\{(P^k, Q^k, R^k)\}_k$  generated by Algorithm 5.1 from any starting point converges to a minimum of the variational problem (5.1).*

**REMARK 5.2.** We remind the readers that the minimizers of the variational principle (5.1) might not be unique. In the non-unique case, the above theorem states that any initial condition will converge to some minimizer, while different initial condition might give different minimizers.

*Proof.* Let  $(P^*, Q^*, R^*)$  be an optimal solution of (5.1). We introduce the short hand notations

$$\begin{aligned} \bar{P}^k &= P^k - P^*, \quad \bar{Q}^k = Q^k - Q^*, \quad \text{and} \quad \bar{R}^k = R^k - R^*. \\ \bar{b}^k &= b^k - b^*, \quad \bar{d}^k = d^k - d^*. \end{aligned} \tag{5.4}$$

From Step 4 and 5 in the algorithm, we get

$$\bar{b}^k = \bar{b}^{k-1} + \bar{P}^k - \bar{Q}^k, \quad \text{and} \quad \bar{d}^k = \bar{d}^{k-1} + \bar{P}^k - \bar{R}^k, \tag{5.5}$$

and hence

$$\begin{aligned} \|\bar{b}^{k-1}\|^2 - \|\bar{b}^k\|^2 &= -2\langle \bar{b}^{k-1}, \bar{P}^k - \bar{Q}^k \rangle - \|\bar{P}^k - \bar{Q}^k\|^2 \\ \|\bar{d}^{k-1}\|^2 - \|\bar{d}^k\|^2 &= -2\langle \bar{d}^{k-1}, \bar{P}^k - \bar{R}^k \rangle - \|\bar{P}^k - \bar{R}^k\|^2. \end{aligned} \tag{5.6}$$

Note that by optimality

$$P^* = \arg \min_{P \in \mathcal{C}_P} \mathcal{L}(P, Q^*, R^*; b^*, d^*) \tag{5.7}$$

$$Q^* = \arg \min_{Q \in \mathcal{C}_Q} \mathcal{L}(P^*, Q, R^*; b^*, d^*) \tag{5.8}$$

$$R^* = \arg \min_{R \in \mathcal{C}_R} \mathcal{L}(P^*, Q^*, R; b^*, d^*). \tag{5.9}$$

Hence, for any  $P, Q, R \in \mathbb{R}^{n \times n}$ , we have

$$f(P) - f(P^*) + \lambda \langle P^* - Q^* + b^*, P - P^* \rangle + r \langle P^* - R^* + d^*, P - P^* \rangle \geq 0 \tag{5.10}$$

$$g(Q) - g(Q^*) + \lambda \langle Q^* - P^* - b^*, Q - Q^* \rangle \geq 0 \tag{5.11}$$

$$h(R) - h(R^*) + r \langle R^* - P^* - d^*, R - R^* \rangle \geq 0. \tag{5.12}$$

According to the construction of  $\{P^k, Q^k, R^k\}$ , for any  $P, Q, R \in \mathbb{R}^{n \times n}$ , we have

$$\begin{aligned} f(P) - f(P^k) + \lambda \langle P^k - Q^{k-1} + b^{k-1}, P - P^k \rangle \\ + r \langle P^k - R^k + d^{k-1}, P - P^k \rangle \geq 0 \end{aligned} \tag{5.13}$$

$$g(Q) - g(Q^k) + \lambda \langle Q^k - P^k - b^{k-1}, Q - Q^k \rangle \geq 0 \tag{5.14}$$

$$h(R) - h(R^k) + r \langle R^k - P^k - d^{k-1}, R - R^k \rangle \geq 0. \tag{5.15}$$

Let  $P = P^k$  in (5.10) and  $P = P^*$  in (5.13), their summation yields

$$\lambda \langle -\bar{P}^k + \bar{Q}^{k-1} - \bar{b}^{k-1}, \bar{P}^k \rangle + r \langle -\bar{P}^k + \bar{R}^{k-1} - \bar{d}^{k-1}, \bar{P}^k \rangle \geq 0. \tag{5.16}$$

Similarly, let  $Q = Q^k$  in (5.11) and  $Q = Q^*$  in (5.14), and let  $R = R^k$  in (5.12) and  $R = R^*$  in (5.15), we obtain

$$\lambda \langle -\bar{Q}^k + \bar{P}^k + \bar{b}^{k-1}, \bar{Q}^k \rangle \geq 0 \tag{5.17}$$

$$r \langle -\bar{R}^k + \bar{P}^k + \bar{d}^{k-1}, \bar{R}^k \rangle \geq 0. \tag{5.18}$$

The summation of (5.16), (5.17), and (5.18) yields

$$\begin{aligned} &\lambda \langle -\bar{b}^{k-1}, \bar{P}^k \rangle + \lambda \langle \bar{Q}^{k-1} - \bar{P}^k, \bar{P}^k \rangle + r \langle -\bar{d}^{k-1}, \bar{P}^k \rangle + r \langle \bar{R}^{k-1} - \bar{P}^k, \bar{P}^k \rangle \\ &\quad + \lambda \langle \bar{b}^{k-1}, \bar{Q}^k \rangle + \lambda \langle \bar{P}^k - \bar{Q}^k, \bar{Q}^k \rangle + r \langle \bar{d}^{k-1}, \bar{R}^k \rangle + r \langle \bar{P}^k - \bar{R}^k, \bar{R}^k \rangle \geq 0. \end{aligned} \tag{5.19}$$

This gives us, after organizing terms

$$\begin{aligned} &-\lambda \langle \bar{b}^{k-1}, \bar{P}^k - \bar{Q}^k \rangle - \lambda \|\bar{Q}^k - \bar{P}^k\|^2 - \lambda \langle \bar{P}^k, \bar{Q}^k - \bar{Q}^{k-1} \rangle \\ &\quad - r \langle \bar{d}^{k-1}, \bar{P}^k - \bar{R}^k \rangle - r \|\bar{R}^k - \bar{P}^k\|^2 - r \langle \bar{P}^k, \bar{R}^k - \bar{R}^{k-1} \rangle \geq 0. \end{aligned} \tag{5.20}$$

Combining the above inequality with (5.6), we have

$$\begin{aligned} &(\lambda \|\bar{b}^{k-1}\|^2 + r \|\bar{d}^{k-1}\|^2) - (\lambda \|\bar{b}^k\|^2 + r \|\bar{d}^k\|^2) \\ &= \lambda (-2 \langle \bar{b}^{k-1}, \bar{P}^k - \bar{Q}^k \rangle - \|\bar{P}^k - \bar{Q}^k\|^2) + r (-2 \langle \bar{d}^{k-1}, \bar{P}^k - \bar{R}^k \rangle - \|\bar{P}^k - \bar{R}^k\|^2) \\ &\geq \lambda \|\bar{Q}^k - \bar{P}^k\|^2 + 2\lambda \langle \bar{P}^k, \bar{Q}^k - \bar{Q}^{k-1} \rangle + r \|\bar{R}^k - \bar{P}^k\|^2 + 2r \langle \bar{P}^k, \bar{R}^k - \bar{R}^{k-1} \rangle. \end{aligned} \tag{5.21}$$

Now, we calculate  $\langle \bar{P}^k, \bar{Q}^k - \bar{Q}^{k-1} \rangle$ . It is clear that

$$\begin{aligned} &\langle \bar{P}^k, \bar{Q}^k - \bar{Q}^{k-1} \rangle \\ &= \langle \bar{P}^k - \bar{P}^{k-1}, \bar{Q}^k - \bar{Q}^{k-1} \rangle + \langle \bar{P}^{k-1} - \bar{Q}^{k-1}, \bar{Q}^k - \bar{Q}^{k-1} \rangle + \langle \bar{Q}^{k-1}, \bar{Q}^k - \bar{Q}^{k-1} \rangle. \end{aligned} \tag{5.22}$$

Note that  $Q^{k-1} = \operatorname{argmin}_Q g(Q) + \frac{\lambda}{2} \|Q - P^{k-1} - b^{k-2}\|^2$ . Thus, for any  $Q \in \mathbb{R}^{n \times n}$ , we have

$$g(Q) - g(Q^{k-1}) + \lambda \langle Q^{k-1} - P^{k-1} - b^{k-2}, Q - Q^{k-1} \rangle \geq 0. \tag{5.23}$$

In particular, let  $Q = Q^k$ , we have

$$g(Q^k) - g(Q^{k-1}) + \lambda \langle Q^{k-1} - P^{k-1} - b^{k-2}, Q^k - Q^{k-1} \rangle \geq 0. \tag{5.24}$$

On the other hand, set  $Q = Q^{k-1}$  in (5.14), we get

$$g(Q^{k-1}) - g(Q^k) + \lambda \langle Q^k - P^k - b^{k-1}, Q^{k-1} - Q^k \rangle \geq 0. \tag{5.25}$$

The summation of (5.24) and (5.25) yields

$$\langle b^{k-1} - b^{k-2}, Q^k - Q^{k-1} \rangle + \langle Q^{k-1} - Q^k + P^k - P^{k-1}, Q^k - Q^{k-1} \rangle \geq 0. \tag{5.26}$$

Note that  $P^k - P^{k-1} = \bar{P}^k - \bar{P}^{k-1}$ ,  $Q^k - Q^{k-1} = \bar{Q}^k - \bar{Q}^{k-1}$ , and  $b^{k-1} - b^{k-2} = \bar{P}^{k-1} - \bar{Q}^{k-1}$ , thus we have

$$\langle \bar{P}^{k-1} - \bar{Q}^{k-1}, \bar{Q}^k - \bar{Q}^{k-1} \rangle + \langle \bar{P}^k - \bar{P}^{k-1}, \bar{Q}^k - \bar{Q}^{k-1} \rangle \geq \| \bar{Q}^k - \bar{Q}^{k-1} \|^2. \tag{5.27}$$

Combine (5.27) with (5.22), we have

$$\langle \bar{P}^k, \bar{Q}^k - \bar{Q}^{k-1} \rangle \geq \| \bar{Q}^k - \bar{Q}^{k-1} \|^2 + \langle \bar{Q}^{k-1}, \bar{Q}^k - \bar{Q}^{k-1} \rangle. \tag{5.28}$$

Similarly, we have

$$\langle \bar{P}^k, \bar{R}^k - \bar{R}^{k-1} \rangle \geq \| \bar{R}^k - \bar{R}^{k-1} \|^2 + \langle \bar{R}^{k-1}, \bar{R}^k - \bar{R}^{k-1} \rangle. \tag{5.29}$$

Substituting (5.28) and (5.29) into (5.21), we have

$$\begin{aligned} & (\lambda \| \bar{b}^{k-1} \|^2 + r \| \bar{d}^{k-1} \|^2) - (\lambda \| \bar{b}^k \|^2 + r \| \bar{d}^k \|^2) \\ & \geq \lambda \| \bar{Q}^k - \bar{P}^k \|^2 + 2\lambda \langle \bar{P}^k, \bar{Q}^k - \bar{Q}^{k-1} \rangle + r \| \bar{R}^k - \bar{P}^k \|^2 + 2r \langle \bar{P}^k, \bar{R}^k - \bar{R}^{k-1} \rangle \\ & \geq \lambda \| \bar{Q}^k - \bar{P}^k \|^2 + 2\lambda (\| \bar{Q}^k - \bar{Q}^{k-1} \|^2 + \langle \bar{Q}^{k-1}, \bar{Q}^k - \bar{Q}^{k-1} \rangle) \\ & \quad + r \| \bar{R}^k - \bar{P}^k \|^2 + 2r (\| \bar{R}^k - \bar{R}^{k-1} \|^2 + \langle \bar{R}^{k-1}, \bar{R}^k - \bar{R}^{k-1} \rangle) \\ & = \lambda \| \bar{Q}^k - \bar{P}^k \|^2 + \lambda (\| \bar{Q}^k \|^2 - \| \bar{Q}^{k-1} \|^2 + \| \bar{Q}^k - \bar{Q}^{k-1} \|^2) \\ & \quad + r \| \bar{R}^k - \bar{P}^k \|^2 + r (\| \bar{R}^k \|^2 - \| \bar{R}^{k-1} \|^2 + \| \bar{R}^k - \bar{R}^{k-1} \|^2) \end{aligned} \tag{5.30}$$

which yields

$$\begin{aligned} & (\lambda \| \bar{b}^{k-1} \|^2 + r \| \bar{d}^{k-1} \|^2 + \lambda \| \bar{Q}^{k-1} \|^2 + r \| \bar{R}^{k-1} \|^2) \\ & \quad - (\lambda \| \bar{b}^k \|^2 + r \| \bar{d}^k \|^2 + \lambda \| \bar{Q}^k \|^2 + r \| \bar{R}^k \|^2) \\ & \geq \lambda \| \bar{Q}^k - \bar{P}^k \|^2 + \lambda \| \bar{Q}^k - \bar{Q}^{k-1} \|^2 + r \| \bar{R}^k - \bar{P}^k \|^2 + r \| \bar{R}^k - \bar{R}^{k-1} \|^2. \end{aligned} \tag{5.31}$$

This concludes that the sequence  $\left\{ \lambda \| \bar{b}^k \|^2 + r \| \bar{d}^k \|^2 + \lambda \| \bar{Q}^k \|^2 + r \| \bar{R}^k \|^2 \right\}_k$  is non-increasing and hence convergent. This further implies,

- (a)  $\{P^k\}_k, \{Q^k\}_k, \{R^k\}_k, \{b^k\}_k$ , and  $\{d^k\}_k$  are all bounded sequences, and hence the sequences has limit points.
- (b)  $\lim_{k \rightarrow \infty} \|Q^k - P^k\| = 0$  and  $\lim_{k \rightarrow \infty} \|R^k - P^k\| = 0$ .

Therefore, the sequences have limit points. Let us denote  $(\tilde{P}, \tilde{Q}, \tilde{R}; \tilde{b}, \tilde{d})$  as a limit point, that is, a subsequence converges

$$\lim_{j \rightarrow \infty} (P^{k_j}, Q^{k_j}, R^{k_j}; b^{k_j}, d^{k_j}) = (\tilde{P}, \tilde{Q}, \tilde{R}; \tilde{b}, \tilde{d}). \tag{5.32}$$

We now prove that  $(\tilde{P}, \tilde{Q}, \tilde{R})$  is a minimum of the variational problem (5.1), *i.e.*

$$f(\tilde{P}) + g(\tilde{Q}) + h(\tilde{R}) = \lim_{j \rightarrow \infty} f(P^{k_j}) + g(Q^{k_j}) + h(R^{k_j}) = f(P^*) + g(Q^*) + h(R^*). \tag{5.33}$$

First note that since  $(P^*, Q^*, R^*; b^*, d^*)$  is a saddle point, we have

$$\begin{aligned} f(P^*) + g(Q^*) + h(R^*) & \leq f(P^{k_j}) + g(Q^{k_j}) + h(R^{k_j}) + \frac{\lambda}{2} \|P^{k_j} - Q^{k_j}\|^2 \\ & \quad + \lambda \langle P^{k_j} - Q^{k_j}, b^* \rangle + \frac{r}{2} \|P^{k_j} - R^{k_j}\|^2 + r \langle P^{k_j} - R^{k_j}, d^* \rangle. \end{aligned} \tag{5.34}$$

Taking the limit  $j \rightarrow \infty$ , we get

$$f(P^*) + g(Q^*) + h(R^*) \leq f(\tilde{P}) + g(\tilde{Q}) + h(\tilde{R}). \quad (5.35)$$

On the other hand, taking  $P = P^*$ ,  $Q = Q^*$ , and  $R = R^*$  in (5.13)–(5.15), we get

$$\begin{aligned} & f(P^*) + g(Q^*) + h(R^*) \\ \geq & f(P^{k_j}) + g(Q^{k_j}) + h(R^{k_j}) - \lambda \langle P^{k_j} - Q^{k_j-1} + b^{k_j-1}, P^* - P^{k_j} \rangle \\ & - r \langle P^{k_j} - R^{k_j} + d^{k_j-1}, P^* - P^{k_j} \rangle - \lambda \langle Q^{k_j} - P^{k_j} - b^{k_j-1}, Q^* - Q^{k_j} \rangle \\ & - r \langle R^{k_j} - P^{k_j} - d^{k_j-1}, R^* - R^{k_j} \rangle \\ = & f(P^{k_j}) + g(Q^{k_j}) + h(R^{k_j}) \\ & - \lambda \langle b^{k_j-1}, Q^{k_j} - P^{k_j} \rangle - \lambda \langle P^{k_j} - Q^{k_j-1}, P^* - P^{k_j} \rangle - \lambda \langle Q^{k_j} - P^{k_j}, Q^* - Q^{k_j} \rangle \\ & - r \langle d^{k_j-1}, R^{k_j} - P^{k_j} \rangle - r \langle P^{k_j} - R^{k_j}, P^* - P^{k_j} \rangle - r \langle R^{k_j} - P^{k_j}, R^* - R^{k_j} \rangle. \end{aligned}$$

From (5.31), we have  $\{P^{k_j}\}$ ,  $\{Q^{k_j}\}$ ,  $\{R^{k_j}\}$ ,  $\{b^{k_j}\}$ , and  $\{d^{k_j}\}$  are all bounded sequences, and furthermore,

$$\begin{aligned} \lim_{j \rightarrow \infty} \|Q^{k_j} - P^{k_j}\| &= 0, & \lim_{j \rightarrow \infty} \|Q^{k_j} - Q^{k_j-1}\| &= 0. \\ \lim_{j \rightarrow \infty} \|R^{k_j} - P^{k_j}\| &= 0, & \lim_{j \rightarrow \infty} \|R^{k_j} - R^{k_j-1}\| &= 0. \end{aligned}$$

Taking the limit  $j \rightarrow \infty$ , we then get

$$f(P^*) + g(Q^*) + h(R^*) \geq f(\tilde{P}) + g(\tilde{Q}) + h(\tilde{R}). \quad (5.36)$$

Hence, the limit point is a minimizer of the variational principle.

Finally, repeating the derivation of (5.31) by replacing  $(P^*, Q^*, R^*)$  by  $(\tilde{P}, \tilde{Q}, \tilde{R})$ , we get convergence of the whole sequence due to the monotonicity.  $\square$

#### REFERENCES

- [1] S Baroni and P. Giannozzi, *Towards very large-scale electronic-structure calculations*, Europhys. Lett., 17, 547–552, 1992.
- [2] E.J. Candes, T. Strohmer, and V. Voroninski, *PhaseLift: exact and stable recovery from magnitude measurements via convex programming*, Commun. Pure Appl. Math., 66, 1241–1274, 2013.
- [3] J. des Cloizeaux, *Analytical properties of n-dimensional energy bands and Wannier functions*, Phys. Rev., 135, A698–A707, 1964.
- [4] J. des Cloizeaux, *Energy bands and projection operators in a crystal: analytic and asymptotic properties*, Phys. Rev., 135, A685–A697, 1964.
- [5] W. E, T. Li, and J. Lu, *Localized bases of eigensubspaces and operator compression*, Proc. Natl. Acad. Sci. USA, 107, 1273–1278, 2010.
- [6] W. E and J. Lu, *The electronic structure of smoothly deformed crystals: Cauchy–Born rule for nonlinear tight-binding model*, Commun. Pure Appl. Math., 63, 1432–1468, 2010.
- [7] W. E and J. Lu, *The Kohn–Sham equation for deformed crystals*, Mem. Amer. Math. Soc., 221, 1040, 2013.
- [8] E. Esser, *Applications of Lagrangian-based alternating direction methods and connections to split Bregman*, UCLA CAM Report, 09–31, 2009.
- [9] C.J. Garcia-Cervera, J. Lu, Y. Xuan, and W. E, *A linear scaling subspace iteration algorithm with optimally localized non-orthogonal wave functions for Kohn–Sham density functional theory*, Phys. Rev. B, 115110, 79, 2009.
- [10] S. Geodecker, *Linear scaling electronic structure methods*, Rev. Mod. Phys., 1085–1123, 71, 1999.
- [11] R. Glowinski and P. Le Tallec, *Augmented Lagrangian and Operator-splitting Methods in Nonlinear Mechanics*, SIAM, Philadelphia, 1989.

- [12] T. Goldstein and S. Osher, *The split Bregman method for  $L_1$ -regularized problems*, SIAM Journal on Imaging Sciences, 2(2), 323–343, 2009.
- [13] S. Kivelson, *Wannier functions in one-dimensional disordered systems: Application to fractionally charged solitons*, Phys. Rev. B, 26, 4269–4277, 1982.
- [14] W. Kohn, *Analytic properties of Bloch waves and Wannier Functions*, Phys. Rev., 115(4), 809–821, 1959.
- [15] W. Kohn, *Analytic properties of Bloch waves and Wannier functions*, Phys. Rev., 115, 809–821, 1959.
- [16] R. de L. Kronig and WG Penney, *Quantum mechanics of electrons in crystal lattices*, Proc. R. Soc. Lond. Ser. A, 130(814), 499–513, 1931.
- [17] R. Lai and S. Osher, *A splitting method for orthogonality constrained problems*, J. Sci. Comput., 58(2), 431–449, 2014.
- [18] X.-P. Li, R.W. Nunes, and D. Vanderbilt, *Density-matrix electronic-structure method with linear system-size scaling*, Phys. Rev. B, 47, 10891–10894, 1993.
- [19] E.H. Lieb and B. Simon, *The Hartree–Fock theory for Coulomb systems*, Commun. Math. Phys., 53, 185–194, 1977.
- [20] Nicola Marzari and David Vanderbilt, *Maximally localized generalized Wannier functions for composite energy bands*, Phys. Rev. B, 56(20), 12847–12865, 1997.
- [21] R. McWeeny, *Some recent advances in density matrix theory*, Rev. Mod. Phys., 32, 335–369, 1960.
- [22] A. Nenciu and G. Nenciu, *The existence of generalised Wannier functions for one-dimensional systems*, Commun. Math. Phys., 190, 541–548, 1998.
- [23] G. Nenciu, *Existence of the exponentially localised Wannier functions*, Commun. Math. Phys., 91, 81–85, 1983.
- [24] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin, *An iterative regularization method for total variation-based image restoration*, Multiscale Model. Simul., 4, 460–489, 2005.
- [25] V. Ozolins, R. Lai, R. Caffisch, and S. Osher, *Compressed modes for variational problems in mathematics and physics*, Prol. Natl. Acad. Sci. USA, 110, 18368–18373, 2013.
- [26] G. Panati, *Triviality of Bloch and Bloch–Dirac bundles*, Ann. Henri Poincaré, 8, 995–1011, 2007.
- [27] S. Setzer, *Split Bregman algorithm, Douglas–Rachford splitting and frame shrinkage*, Proceedings of the 2nd International Conference on Scale Space and Variational Methods in Computer Vision, LNCS, 5567, 2009.
- [28] G.H. Wannier, *The structure of electronic excitation levels in insulating crystals*, Phys. Rev., 52(3), 0191–0197, 1973-08.
- [29] C. Wu and X. Tai, *Augmented Lagrangian method, dual methods and split-Bregman iterations for ROF, vectorial TV and higher order models*, SIAM J. Imaging Sci., 3(3), 300–339, 2010.
- [30] W. Yin and S. Osher, *Error forgetting of Bregman iteration*, J. Sci. Comput., 54(2-3), 684–695, 2013.
- [31] W. Yin, S. Osher, D. Goldfarb, and J. Darbon, *Bregman iterative algorithms for  $\ell_1$ -minimization with applications to compressed sensing*, SIAM J. on Imaging Sci., 1(1), 143–168, 2008.