This is a 'slow' derivation of the SpM method by Shinaoka and collaborators

June 1, 2017

1 Analytic continuation equations

We start with the analytic continuation problem

$$G = K\rho \tag{1}$$

where G denotes the input data, usually in imaginary time or frequency, K denotes the analytic continuation kernel, and ρ denotes the desired spectral function multiplied with a frequency element $\Delta\omega$.

K is difficult to invert but elementary linear algebra says that we can decompose it into

$$K = U\Sigma V^T \tag{2}$$

and that the matrix Σ can be compressed suitably by removing eigenvalues that are zero to within numerical precision.

If we approximate ρ rather than solve it exactly, we can characterize the deviation between some approximate ρ and the original input data as

$$\chi^2 = ||G - K\rho||_2^2. \tag{3}$$

We will use this norm to characterize deviations from original data.

The premise of many analytic continuation methods is that optimizing (or solving) for $\chi^2 = 0$ is not a good idea due to the bad conditioning of K. Rather, one wants to find a solution which minimizes χ^2 under some additional constraint condition. SpM is one of these methods, and the constraint is formulated based on the SVD described above:

$$\chi^2 = ||G - K\rho||_2^2 = ||U^T G - \Sigma V^T \rho||_2^2 =: ||y' - Sx'||_2^2 \tag{4}$$

where we define $S = \Sigma$, $y' = U^T G$, and $x' = V^T \rho$. Shinaoka and collaborators postulate that it is a good idea to minimize

$$|x'|_1 = \sum_j |x'_j|, \tag{5}$$

and therefore one should minimize

$$L_\lambda(x') = \frac{1}{\lambda}||y' - Sx'||_2^2 + |x'|_1, \tag{6}$$

where λ is an (a priori arbitrary) parameter used to penalize larger x'. The L_1 norm is chosen because the minimization with that constraint is easily doable using ADMM but seems to have no other justification.

In addition, we would like to satisfy the constraint that ρ is a decent spectral function. That means that we'd like to guarantee that it integrate to one, and that it is exclusively positive. Thus, the whole problem is: Minimize $L_\lambda(x')$ under the constraints

$$\sum_j \rho_j = \sum_j V_{jk} x'_k = 1$$

$$\rho_j = V x'_j > 0 \;\forall j \tag{7}$$

Note that the way the Maxent code was set up, ρ already contains a multiplication by a frequency element, such that the spectral function at frequency index k is given by $\rho_k/\Delta\omega_k$. The optimization problem is of the form of a constrained LASSO problem with two constraints: a constraint that requires a convex optimization (the positivity), and a constraint for the norm. For more possible constraints see below.

2 Optimization without constraints

The optimization without constraints of Eq.6 is fairly straightforward with ADMM: we follow section 6.4, on LASSO, and minimize

$$L^\lambda(x') = \frac{1}{2}||Sx' - y'||_2^2 + \lambda|x'|_1 \tag{8}$$

which we write as

$$\text{minimize} \;\; f(x') + g(z') \tag{9}$$

$$\text{subject to} \;\; x' - z' = 0 \tag{10}$$

where we define $f(x') = \frac{1}{2}||Sx' - y'||_2^2$ and $g(z') = \lambda|z'|_1$. The optimization procedure then consists of three separate steps:

$$x'^{k+1} := (S^T S + \rho I)^{-1}(S^T y' + \rho(z'^k - u^k)) \tag{11}$$

$$z'^{k+1} = S_{(\lambda/\rho)}(x'^{k+1} + u^k) \tag{12}$$

$$u^{k+1} = u^k + x'^{k+1} - z'^{k+1} \tag{13}$$

where we've introduced a scalar parameter ρ, a vector u, and a soft thresholding function S. These equations implement ADMM in the so-called scaled form, and the converged solution will not depend on ρ. [Note that ρ here is not the spectral function ρ introduced earlier on pg.1.]

3 Scaled form of ADMM

We repeat the essentials of the derivation here, following chapter 3. The ADMM algorithm solves optimization problems of the form

$$\text{minimize} \quad f(x) + g(z) \tag{14}$$

$$\text{subject to} \quad Ax + Bz = c \tag{15}$$

by forming an augmented lagrangian with Lagrangian parameter ρ

$$L_\rho(x, z, y) = f(x) + g(z) + y^T(Ax + Bz - c) + \frac{\rho}{2}||Ax + Bz - c||_2^2 \tag{16}$$

The first part would be the Lagrangian without the 'augmented' constraint. The second part will be zero if the constraint is satisfied. ADMM then does the iterations

$$x^{k+1} = \text{argmin}_x L_\rho(x, z^k, y^k) \tag{17}$$

$$z^{k+1} = \text{argmin}_z L_\rho(x^{k+1}, z, y^k) \tag{18}$$

$$y^{k+1} = y^k + \rho(Ax^{k+1} + Bz^{k+1} - c) \tag{19}$$

This is equivalent to the so-called scaled form, which scales the dual variable y as $u = \frac{1}{\rho}y$, such that

$$x^{k+1} = \text{argmin}_x \left(f(x) + \frac{\rho}{2}||Ax + Bz^k - c + u^k||_2^2 \right) \tag{20}$$

$$z^{k+1} = \text{argmin}_z \left(g(z) + \frac{\rho}{2}||Ax^{k+1} + Bz - c + u^k||_2^2 \right) \tag{21}$$

$$u^{k+1} = u^k + Ax^{k+1} + Bz^{k+1} - c \tag{22}$$

The vector u^k sums up the residuals $r^k = Ax^k + Bz^k - c$, such that $u^k = u^0 + \sum_{j=1}^k r^j$.

4 Optimization in x

We start with the description of the optimization in x, which is $x'^{k+1} = \text{argmin}_{x'}(f(x') + \frac{\rho}{2}||Ax' + Bz'^k - c + u^k||_2^2)$ using $A = I$, $B = -I$, $c = 0$ or, equivalently, the constraint $x' - z' = 0$. This implies

$$x'^{k+1} = \text{argmin}_{x'} \left(\frac{1}{2}||Sx' - y'||_2^2 + \frac{\rho}{2}||x' - z'^k + u^k||_2^2 \right) \tag{23}$$

Let us multiply this out:

$$\frac{1}{2}\left[(Sx' - y')^T(Sx' - y') + \rho(x' - (z'^k - u^k))^T(x' - (z'^k - u^k))\right] = \frac{1}{2}x'^T Ax' - x'^T b + const \tag{24}$$

where

$$A = S^T S + \rho I \tag{25}$$

$$b = S^T(y') + \rho(z'^k - u^k) \tag{26}$$

which has a minimum exactly where $Ax = b$, or

$$(S^T S + \rho I)x'^+ = S^T(y') + \rho(z'^k - u^k). \tag{27}$$

where $v = z'^k - u^k$ (note that we don't need the constant term to minimize this – why introduce v here???).

Plugging these expressions back into Eq. 23, and writing an explicit inverse rather than an equation solution for x'^{k+1} we obtain

$$x'^{k+1} = (S^T S + \rho I)^{-1}(S^T(y') + \rho(z'^k - u^k)). \tag{28}$$

5 Optimization in z

The second optimization is the optimization for the L_1 term. It comes from the optimization of

$$z'^{k+1} = \text{argmin}_{z'}\left(g(z') + \frac{\rho}{2}||x'^{k+1} - z' + u^k||_2^2\right) = \text{argmin}_{z'}\left(\lambda|z'|_1 + \frac{\rho}{2}||z' - (x'^{k+1} + u^k)||_2^2\right) \tag{29}$$

where $g(z') = \lambda|z'|_1$ and, defining $v = x'^{k+1} + u^k$

$$z'^{k+1} = \text{argmin}_{z'}\left(\lambda|z'|_1 + \frac{\rho}{2}||z' - v||_2^2\right) \tag{30}$$

The equation above does not mix different elements of the vector z'. Thus we can rewrite it as an element-wise minimization and do the update for every element of z'individually, such that

$$z_j'^{k+1} = \text{argmin}_{z_j'}\left(\lambda|z_j'|_1 + \frac{\rho}{2}(z_j' - v_j)^2\right) \tag{31}$$

The solution to this minimization is given by the soft thresholding operator

$$z_j'^{+} = S_{\lambda/\rho}(v_j) \tag{32}$$

where the soft thresholding operator S is given by

$$S_\kappa(x) = \begin{cases} x - \kappa & x > \kappa \\ 0 & |x| \leq \kappa \\ x + \kappa & x < -\kappa \end{cases} \tag{33}$$

This is derived in detail in chapter 4.4.3.

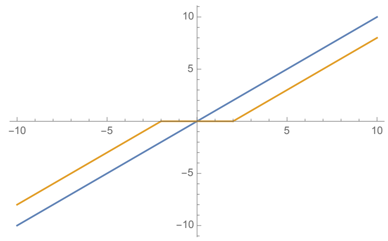

Figure 1:
This is the soft thresholding function applied to x (in blue) for a threshold of 2. Everything moves closer to zero.

6 The soft thresholding function in more detail

Here we derive the formula for $S_\kappa(x)$ in more detail. We start by defining the problem as

$$prox_f(x) = \text{argmin}_z(\frac{\rho}{2}||z-v||_2^2 + \lambda|z|_1) \tag{34}$$

which, because it is separable, we can write as

$$\frac{\partial}{\partial z}(\frac{\rho}{2}(z_j - v_j)^2 + \lambda|z_j|) \stackrel{!}{=} 0 \tag{35}$$

which is

$$z_j + \frac{\lambda}{\rho}\frac{\partial}{\partial z_j}|z_j| = v_j \tag{36}$$

if $|z_j| > 0$, the derivative is 1, if $|z_j| < 0$, it is -1, and if $|z_j| = 0$ we can't really differentiate. Let's do the case $z > 0$ first:

$$z_j = v - \frac{\lambda}{\rho}. \tag{37}$$

Otherwise we pick up a -1, and the equation is

$$z_j = v + \frac{\lambda}{\rho} \tag{38}$$

Finally let's examine the case of $|z_j| = 0$. The literature works with a subdifferential and states $z_j + \frac{\lambda}{\rho}[-1, 1] = 0$, which has a solution of 0 for $z_j$in $[-\frac{\lambda}{\rho}, \frac{\lambda}{\rho}]$. Putting it all together, we get

$$z_i = \begin{cases} 0 & |v_i| < \frac{\lambda}{\rho} \\ v_i - \frac{\lambda}{\rho} sgn(x_i) & |v_i| > \frac{\lambda}{\rho} \end{cases} \tag{39}$$

This is illustrated in Fig.1. The equations get mildly more interesting if we minimize

$$argmin_z \left(\frac{\rho}{2}||Ax^{k+1} + Bz - c + u^k||_2^2 + \lambda|z|_1\right) \tag{40}$$

we then define

$$w = Ax^{k+1} + u^k - c \tag{41}$$

and minimize

$$argmin_z \left(\frac{\rho}{2}||Bz + w||_2^2 + \lambda|z|_1\right) = argmin_z \left(\frac{\rho}{2}(z^T B^T Bz + 2z^T B^T w + w^T w) + \lambda|z|_1\right) \tag{42}$$

We are only interested here in the case of $B^T B = 1$, which allows us to redefine $v = -B^T w$ to express the problem as the minimization problem

$$argmin_z \left(\frac{\rho}{2}||z - v||_2^2 + \lambda|z|_1\right) \tag{43}$$

7 The modification for a positivity constraint

We want to enforce positivity of a spectral function. That means that we want to constrain the fit of x' or z' in such a way that the resulting spectral function is positive everywhere. In our notation the spectral function is $\rho = Vx'$. We build in this constraint by modifying the Lagrangian from $L_\rho(x', z')$ to

$$L_\rho(x', z', z) = f(x') + g(z') + h(z) \tag{44}$$

i.e. we introduce a third variable and alternate the optimization between the three of them, subject to the constraint that $x' = z' = V^T z$.

In the textbook case, we just have to minimize $f(x)+h(z)$. In that case, $h(z)$ is constructed such that it enforces a positivity constraint such as in chapter 5, first page, and we obtain

$$z^{k+1} = \Pi_+(x^{k+1} + u^k) \tag{45}$$

The operator Π_+ is a projection operator onto the domain $\mathcal{C}$ where we want to enforce the values of z, i.e. here onto the positive domain. To be concrete, we minimize

$$z^{k+1} = \text{argmin}_z(h(z) + \frac{\rho}{2}||x^{k+1} - z + u^k||_2^2) \tag{46}$$

with $h(z)$ zero only in $\mathcal{C}$ (we can model it as being infinity outside of $\mathcal{C}$). The solution to this is the point where the $||x^{k+1} - z + u^k||$ is smallest with z in the domain $\mathcal{C}$. In our case of a positivity constraint, the function is minimized for

$$z_j = \begin{cases} 0 & (x^{k+1} + u^k) < 0 \\ x^{k+1} + u^k & \text{else} \end{cases} \tag{47}$$

This is also called an Euclidian projection.

Back to our case: In the Lagrangian Eq. 44, $h(z)$ evaluates to zero if z is positive, and to inifinity otherwise. We optimize the three constraint in alternating fashion, in two ADMM loops which independently take care of the optimization of z' and of z: first optimize x' keeping z' and z constant. Then optimize z' keeping x' constant, and update the running sum of residual for this optimization, which is called u'. Independently optimize z keeping x' constant. Then update the running residual sum u for that optimization.

This may become clearer if we relabel z as s and work with $s' = V^T s$, $s = Vs'$ in singular space. In that case

$$L_\rho(x', z', s') = f(x') + g(z') + j(s') \tag{48}$$

and $j(s') = h(Vs')$ is the indicator function evaluated after a transform from singular to direct space. The constraint then is that $x' = z' = s'$. Denoting the running residual vector of the optimization in s' as w', we obtain

$$s'^{k+1} = V^T \Pi_+(Vx'^{k+1} + Vw'^k) \tag{49}$$

which is satisfied by $s' = V^T s$ and

$$s_j = \begin{cases} 0 & Vx'^{k+1} + Vw'^k < 0 \\ Vx'^{k+1} + Vw'^k & \text{else} \end{cases} \tag{50}$$

After the optimization of s' we need to update the residual vector as

$$w'^{k+1} = w'^k + x'^{k+1} - s'^{k+1} \tag{51}$$

which, because it is linear, we can also write in direct space as

$$w^{k+1} = w^k + x^{k+1} - s^{k+1} \tag{52}$$

Formulating this constraint in direct space rather than in singular space makes the equations look easier, which is why they probably did it in the paper. The only consequence of optimizing multiple constraints is having multiple residuals in the equation for x' :

$$x'^{k+1} = \text{argmin}_x \left(\frac{1}{2}||Sx' - y'||_2^2 + \frac{\rho}{2}||x' - z'^k + u^k||_2^2 + \frac{\rho'}{2}||x' - s'^k + w^k|| \right) \tag{53}$$

multiplying all of this out will lead to additional terms with $s'$and w wherever we had terms with z' and u, such that the linear term in the equation for the optimization get modified to

$$b = S^T(y') + \rho(z'^k - u^k) + \rho'(s'^k - w^k) \tag{54}$$

The quadratic terms picks up a $I(\rho + \rho')$ where we had an $I\rho$, and the constant term is irrelevant for the optimization.

8 The modification for a norm constraint

DEPRECATED. I THINK THIS IS CLUMSY.

In order to have spectral functions with a fixed norm of one, we need to optimize under the constraint

$$\sum_j V_{ji} x'_i = 1 \tag{55}$$

We do this in just the same way: modify the Lagrangian to

$$L_\rho(x', z', s', t') = f(x') + g(z') + j(s') + k(t') \tag{56}$$

where $k(t') = \nu(\langle V t' \rangle - 1)$ and, as in the paper, $\langle \cdot \rangle$denotes the integral over the spectral function, which we can also write as $\langle x \rangle = e \cdot x$, where the vector $e$is a unit vector chosen such that $\int d\omega A(\omega) \approx \sum_j x_j$. As a consequence, we pick up yet another optimization step and yet another vector of residuals v', such that

$$x'^{k+1} = \mathrm{argmin}_x (\frac{1}{2}||Sx' - y'^k||_2^2 + \frac{\rho}{2}||x' - z'^k + u^k||_2^2 + \frac{\rho'}{2}||x' - s'^k + w'^k|| + \frac{\rho''}{2}||x' - t'^k + v'^k||) \tag{57}$$

this will modify the quadratic part of the optimization to

$$A = S^T S + (\rho + \rho' + \rho'')I \tag{58}$$

and change the linear part to

$$b = S^T(y'^k) - \rho(z'^k - u^k) - \rho'(w'^k - s'^k) - \rho''(v'^k - t'^k) \tag{59}$$

and it will require an additional step for the optimization of t and an update of the residuals v. The t update is given as

$$t'^{k+1} = \mathrm{argmin}_{t'}(k(t') - \frac{\rho''}{2}||x'^{k+1} - t' + v'^k||_2^2) \tag{60}$$

and we can write this as the optmization of a quadratic equation, similar to the optimization of x', and obtain the quadratic and linear terms

$$t'^{k+1} = \mathrm{argmin}_{t'} \frac{-\rho''}{2} t'^T t' - (\rho''(x'^{k+1} + {v'}^k)^T + \nu e V)t' + const \tag{61}$$

which results in

$$t'^{k+1} = (x' + v'^k) + \frac{\nu}{\rho''} eV \tag{62}$$

this needs to be followed by an update of the running residuals,

$$v'^{k+1} = v'^k + x'^{k+1} - t'^k \tag{63}$$

9 The modification for high frequency constraints

This is not in the paper, but it is a straightforward extension of the constraints in the paper. We have additional constraints for the spectral functions, which are

$$c_2 = \int \omega A(\omega) \tag{64}$$

$$c_3 = \int \omega^2 A(\omega) \tag{65}$$

These constraints come straight out of the series expansion of $G(i\omega_n) = \frac{1}{i\omega_n} + \frac{c_2}{(i\omega_n)^2} + \frac{c_3}{(i\omega_n)^3} + ...$ and the evaluation of the kernel.

$$n = \int A(\omega) f_\beta(\omega - \mu) \tag{66}$$

This is just the expression for the particle number. Frequently we know these terms analytically, or we at least know then to much higher precision than the spectral function. They can be enforced using a constraint just as the norm constraint.

10 Building in norm and integral constraints exactly

ADMM has a straightforward way of building in constraints: currently we use $x' = z'$, or $Ax' + Bz' - c = 0$ with $A = I$ and $B = -I$ and $c = 0$. A, B, and c do not need to be $n_s \times n_s$ matrices and it should be easy to extend A, B, and c by a row each, such that

$$A = \begin{pmatrix} I_{n_s} \\ \sum_j V_{jk} \end{pmatrix} \tag{67}$$

$$B = \begin{pmatrix} -I_{n_s} \\ 0 \end{pmatrix} \tag{68}$$

$$c = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{69}$$

such that the last line enforces

$$\sum_j V_{jk} x'_k = 1 \tag{70}$$

which is the condition for the norm. This will leave the L_1 optimization of z' invariant but will change the optimization for x' to

$$x'^{k+1} = \text{argmin}_{x'} \left(\frac{1}{2} ||Sx' - y'||_2^2 + \frac{\rho}{2} ||Ax' + Bz'^k - c + u^k||_2^2 \right) \tag{71}$$

Multiplying this out, we get a quadratic equation of the form $\frac{1}{2} x^T P x + q^T x + c = 0$, with a solution at $Px = q$, for

$$P = S^T S + \rho A^T A \tag{72}$$

$$q = y'^T S + \rho (Bz'^k - c + u^k)^T A \tag{73}$$

Additional constraints for $\int d\omega \omega A(\omega)$ and $\int d\omega \omega^2 A(\omega)$ result in additional lines with $\sum_j \omega_j V_{jk}$ and $\sum_j \omega_j^2 V_{jk}$ for A and c_2 and c_3 for c.

11 Summary of Equations

This section summarizes the equations needed to solve the ADMM problem. We start with the calculation of A and B as in 67 and the optimization of x^{k+1} according to Eq.72 and 73.

$$P = S^T S + \rho A^T A \tag{74}$$

$$q = y'^T S + \rho(Bz'^k - c + u^k)^T A - \rho'(w'^k - s'^k) - \rho''(v'^k - t'^k) \tag{75}$$

$$Ax'^{k+1} = b \tag{76}$$

As the matrix A is diagonal, we can replace the last equation by an element-wise division of b by A. We then update the L1 norm and its dual variable according to Eq. 32 and Eq. 33. This will take care of z' and its running residual u.

$$z'^{k+1} = S_{\lambda/\rho}(x'^{k+1} + u^k) \tag{77}$$

$$u^{k+1} = u^k + Ax'^{k+1} + Bz'^{k+1} - c \tag{78}$$

We then satisfy all other constraints, starting with the positivity, variables s' and w':

$$s'^{k+1} = V^T \Pi_+(Vx'^{k+1} + Vw'^k) \tag{79}$$

$$w'^{k+1} = w'^k + x'^{k+1} - s'^{k+1} \tag{80}$$